DEFENCE **R&D** DÉFENSE

An Operational Implementation of a CBRN Sensor-Driven Modeling Paradigm for Stochastic Event Reconstruction

Technical Description of an Advanced Modeling Component of CRTI Project 07-0196TD

E. Yee
DRDC Suffield

Canadä

An Operational Implementation of a CBRN Sensor-Driven Modeling Paradigm for Stochastic Event Reconstruction

Technical Description of an Advanced Modeling Component of CRTI Project 07-0196TD

E. Yee
Defence R&D Canada – Suffield

Principal Author

Original signed by E. Yee

E. Yee

Approved by

Original signed by Mr C. Laforce

Mr C. Laforce
Acting Head, Personal Protection Sector

Approved for release by

Original signed by Dr. R.G. Clewley

Dr R. G. Clewley
Acting Chair, DRDC Suffield DRP

Abstract

This report provides a technical description of the module urbanSOURCE, which is an operational implementation of an innovative sensor-driven modeling paradigm for source reconstruction. This module permits the rapid and robust estimation of the parameters of an unknown source, using a finite number of noisy concentration measurements obtained from a sensor array. The problem is solved using a Bayesian probabilistic inferential framework in which Bayesian probability theory is used to formulate the posterior distribution for the source parameters. Three different model equations have been formulated for the likelihood function, leading to three different models for the posterior distribution of the source parameters. The application of the methodology implemented in urbanSOURCE is illustrated using real dispersion data obtained from two examples (Joint Urban 2003 field experiment in Oklahoma City and European Tracer Experiment) involving contaminant dispersion in highly disturbed flows over urban and complex environments, where the idealizations of horizontal homogeneity and/or temporal stationarity in the flow cannot be applied to simplify the problem.

Résumé

Le présent rapport donne une description technique du module urbanSOURCE, lequel est une mise en application opérationnelle d'un paradigme novateur de modélisation déterminée par des capteurs pour la reconstruction de sources. Ce module permet d'obtenir une évaluation rapide et robuste des paramètres d'une source inconnue, au moyen d'un nombre déterminé de mesures de concentration de bruit obtenues à l'aide d'un réseau de capteurs. Le problème est résolu grâce à l'utilisation d'un cadre inférentiel de probabilités bayésiennes dans lequel la théorie des probabilités bayésiennes est utilisée afin d'établir la distribution a posteriori pour les paramètres sources. Trois équations de modèles ont été formulées pour la fonction de vraisemblance, et se sont soldées par trois modèles pour la distribution a posteriori des paramètres sources. L'application de la méthode mise en oeuvre dans le module urbanSOURCE est illustrée à l'aide des données de dispersion réelles tirées de deux exemples (expérience sur le terrain Joint Urban 2003 à Oklahoma City et expérience européenne sur les traceurs [European Tracer Experiment]) portant sur la dispersion de contaminants dans des écoulements très perturbés en milieux urbains complexes, dans lesquels les idéalisations de l'homogénéité horizontale et/ou de la stationnarité temporelle dans l'coulement ne peuvent pas être utilisées pour simplifier le problème.

Executive summary

An Operational Implementation of a CBRN Sensor-Driven Modeling Paradigm for Stochastic Event Reconstruction

Background: Atmospheric dispersion modeling is important to public security as it provides emergency managers and responders with predictions for plume direction, coverage and lethality required to direct efforts for managing the consequences of a toxic agent release. However, before plume dispersion modeling can be applied, a knowledge of the characteristics of the toxic release (e.g., location, emission rate, time of release) is required. A critical capability gap in current emergency and retrospective management efforts, directed at terrorist incidents involving the clandestine release of a chemical, biological, radiological or nuclear (CBRN) agent into the atmosphere, is the ability to determine the characteristics of the unknown source following the detection of the event by a network of CBRN sensors (source reconstruction problem).

Principal results: Research conducted over the past four years on addressing the difficult source reconstruction problem has matured to the point that the results can be transitioned into an operational capability that can be inserted into a CBRN battlespace. In consequence, an operational implementation of an innovative sensor-driven modeling paradigm for source reconstruction has been realized in the form of a software module urbanSOURCE. This module permits the rapid and robust estimation of the parameters of an unknown source, using a finite number of noisy concentration measurements obtained from a CBRN sensor array. The methodology implemented in urbanSOURCE has been successfully validated using concentration data measured in a real urban environment (Joint Urban 2003 field experiment conducted in Oklahoma City, Oklahoma) involving transport and dispersion of an agent on an urban-industrial complex scale and in a complex terrain environment (European Tracer Experiment) involving transport and dispersion of an agent on a continental scale.

Significance of results: The operational implementation of a source reconstruction capability in urbanSOURCE will enable the rapid estimation of an unknown source term associated with a covert (clandestine) release of a CBRN agent, using the available concentration data measured in real-time by a sensor network, followed by an accurate and timely prediction of the agent's spread and deposition required for making more informed decisions for mitigation of the consequences of the toxic release. The capability provided by urbanSOURCE provides the inextricable linkage of CBRN sensor data with advanced models for atmospheric dispersion (and, more particularly, for urban dispersion), leading potentially to significant improvements in the situational awareness in the battlespace.

Future work: The next step is to integrate urbanSOURCE as an operational capability for source reconstruction into the integrative multiscale urban modeling system implemented in

the computational infrastructure at a government operations facility (Environmental Emergency Response Section at Canadian Meteorological Centre). To complete the sensor-driven modeling paradigm, the system needs to be interfaced with information (warning and reporting) systems for automated data acquisition from CBRN sensors. The incorporation of this capability into a government operations facility will give it the key-enabling tools to provide a 'whole-of-government' (comprehensive) single authoritative source for expert quality-assured sensor-driven CBRN hazard predictions and concomitant decision-support aids, which will form the basis for making significantly improved decisions for responding to and mitigating hazardous release incidents. These products can be used by emergency managers, planners and first responders (civil and military) in various federal, provincial and municipal agencies for informed response decision making in both domestic and international operations, as well as for support of major events of national and international significance [e.g., Vancouver Winter Olympics, Group of Eight (G8) and Twenty (G20) Summits, Francophonie Summit].

Sommaire

An Operational Implementation of a CBRN Sensor-Driven Modeling Paradigm for Stochastic Event Reconstruction

E. Yee ; DRDC Suffield TR 2010-070 ; R & D pour la défense Canada – Suffield ; mai 2010.

Contexte : La modélisation de la dispersion atmosphèrique est importante pour la sécurité publique, car elle donne aux gestionnaires des mesures d'urgence et aux intervenants d'urgence des prévisions sur la direction, la couverture et la létalité des panaches, prévisions nécessaires pour orienter la gestion des conséquences d'un rejet d'agents toxiques. Toutefois, avant d'appliquer la modélisation de la dispersion de panache, il faut connaître les caractéristiques du rejet toxique (p. ex., l'emplacement, le taux d'émissions et l'heure du rejet). Une lacune très importante en matière de capacité liée aux efforts actuels de gestion rétrospective et d'urgences, visant les incidents terroristes associés au rejet clandestin dans l'atmosphère d'agents chimiques, biologiques, radiologiques ou nucléaires (CBRN), est la capacité de déterminer les caractéristiques d'une source inconnue après la détection d'un événement par un réseau de capteurs CBRN (problème de reconstruction de la source).

Résultats principaux : Les recherches effectuées au cours des quatre dernières années sur la résolution du problème difficile de reconstruction de sources ont évolué au point où il est possible de convertir les résultats en une capacité opérationnelle, laquelle peut être insérée dans un espace de bataille CBRN. En conséquence, une mise en application opérationnelle d'un paradigme novateur de modélisation déterminée par des capteurs pour la reconstruction de sources a été réalisée sous la forme du module logiciel urbanSOURCE. Ce module permet d'obtenir une évaluation rapide et robuste des paramètres d'une source inconnue, au moyen d'un nombre déterminé de mesures de concentration de bruit obtenues à l'aide d'un réseau de capteurs. La méthode mise en oeuvre dans le module urbanSOURCE a été validée avec succés á l'aide des données de concentration mesurées dans un environnement urbain réel (expérience sur le terrain Joint Urban 2003 effectuée à Oklahoma City, Oklahoma), qui comprend le transport et la dispersion d'un agent à une échelle complexe urbaine industrielle, et un environnement de terrain complexe (European Tracer Experiment), qui comprend sur le transport et la dispersion d'un agent à l'échelle continentale.

Portée des résultats : La mise en application opérationnelle d'une capacité de reconstruction de sources dans le module urbanSOURCE permettra d'évaluer rapidement un terme source inconnu associé à un rejet secret (clandestin) d'un agent CBRN, à l'aide des données disponibles sur les concentrations mesurées en temps réel par un réseau de capteurs, suivie d'une prédiction précise et au bon moment de la dispersion et du dépôt d'agents qui est nécessaire à la prise de décisions mieux éclairées en vue d'atténuer les conséquences des rejets toxiques. La capacité fournie par le module urbanSOURCE permet l'obtention de liens inextricables entre les données des capteurs CBRN et celles des modèles avancés

pour la dispersion atmosphérique (et, en particulier, pour la dispersion urbaine), et pourrait améliorer grandement la connaissance de la situation dans l'espace de bataille.

Perspectives d'avenir : La prochaine étape consiste à intégrer le module urbanSOURCE, en tant que capacité opérationnelle pour la reconstruction de sources, au système de modélisation urbaine multi échelle intégré mis en oeuvre dans l'infrastructure informatique d'une installation dédiée aux opérations gouvernementales (Division de la réponse aux urgences environnementales du Centre météorologique canadien). Pour compléter le paradigme de modélisation déterminée par des capteurs, il faut interfacer le système avec les systèmes d'informations (avertissement et signalement) aux fins d'acquisition automatisée de données provenant des capteurs CBRN. L'intégration de cette capacité à une installation dédiée aux opérations gouvernementales donnera à cette dernière les principaux outils d'appui en vue de fournir une source unique autorisée " pour l'ensemble du gouvernement " (exhaustive) pour les prévisions de dangers CBRN déterminées par des capteurs dont la qualité est validée par des experts et les aides concomitantes à la décision, qui formeront la base des améliorations importantes à la prise de décisions en vue de répondre aux incidents de dispersion de produits dangereux et de les limiter. Ces produits peuvent être utilisés par les gestionnaires des mesures d'urgence, les planificateurs et les premiers répondants (civils et militaires) de divers organismes fédéraux, provinciaux et municipaux aux fins de prise de décisions éclairées tant dans les opérations nationales que les opérations internationales, ainsi que l'appui des principaux événements qui ont une importance sur le plan national et international (p. ex. : Jeux olympiques d'hiver de Vancouver, le Sommet du Groupe des huit [G8], le Sommet du Groupe des vingt [G20] et le Sommet de la Francophonie).

Table of contents

List of figures

List of tables

1 Introduction

1.1 Background

The environmental and toxicological impact of the mean transport and turbulent diffusion of contaminants released into the atmosphere has become increasingly important in recent years. Considerable interest has been focused on the prediction of mean concentration levels downwind of contaminant sources in the turbulent atmospheric boundary layer. Consequently, atmospheric transport and diffusion models have played an important role in emergency response systems for toxic releases and have been used in calculating the transport, diffusion, and deposition of hazardous chemical, biological, radiological or nuclear (CBRN) materials released (either accidentally or deliberately) into the turbulent atmospheric boundary layer over relatively smooth and horizontally homogeneous surfaces.

Military and civilian (government and commercial) emergency response models commonly use standard Gaussian plume or puff models [1], which employ semi-empirical relationships for plume or puff growth with the mean wind and turbulence fields obtained either from similarity theory or from the use of simple diagnostic wind fields constructed from the interpolation and/or extrapolation of sparse observational data. The advantages of these approaches for wind flow specification are their simplicity, general applicability in simple atmospheric conditions, and, most importantly, their limited computational requirements. While this approach is useful for a landscape that is relatively flat and unobstructed, it is wholly inadequate for surface-atmosphere interactions over "complex" surfaces (viz., most of the real world) such as cities and other built-up areas.

As the fraction of the world's population that lives in cities grows, it is becoming increasingly important to address the urgent problem of the assessment of hazards caused by the release of potentially harmful materials into the urban environment. Should such a release of a noxious contaminant occur—perhaps as a result of a deliberate release using a CBRN agent or an accidental spillage of a toxic industrial material—it is important to be able to predict the transport and dispersion of the plume or cloud of potentially hazardous material as it evolves in the urban canopy where human habitation is concentrated.

It should be noted that the development of models for urban dispersion is very complicated owing to the fact that the urban environment is characterized by extremely diverse length and time scales and complex geometries and interfaces. In particular, a typical urban canopy consists of a large collection of buildings and other obstacles (e.g., cars lining a street, treed areas in city green spaces, etc.) that are aggregated into complex structures. When this rough surface interacts with the atmospheric flow within and above it, the highly disturbed flow field can become extremely complex, exhibiting various flow features such as curved mean streamlines, large velocity gradients, sharp velocity discontinuities, flow separations and reattachments, cavity regions, recirculation zones, and strongly inhomogeneous turbulence. Understanding the complex flow of the wind through and above the urban environment and the dispersion of contaminants released into that flow is both necessary and important. In view of this, we require physically-based urban wind models that can predict the complex spatial-temporal pattern of urban wind statistics required to

"drive" models for the dispersion of contaminants within the street network of an urban environment (where it is venting of street canyons that is important for determination of the contaminant concentrations).

The United States Government Accountability Office (GAO) recently conducted a review of various models used by federal agencies to predict the transport and dispersion of terrorist-related and accidental releases of CBRN materials in urban areas [2]. Based on that review, it was concluded that "evaluations and field testing of plume models developed for urban areas show variable predictions in urban environments" and that "federal agencies' models to track the atmospheric release of CBRN materials have major limitations in urban areas". In addition to these deficiencies, the GAO report also cautioned the reader that using predictions of non-urban plume models for CBRN events in urban areas "are limited in their ability to accurately predict the path of a plume and the extent of contamination in urban environments".

This identified capability gap (which has been generally acknowledged by various dispersion modeling experts) was the motivation for the development of an advanced emergency response system for CBRN hazard prediction and assessment for the urban environment sponsored by **C**hemical, Biological, Radiological-Nuclear and Explosives **R**esearch and **T**echnology **I**nitiative (CRTI) under Project 02-0093RD entitled "An Advanced Emergency Response System for CBRN Hazard Prediction and Assessment for the Urban Environment". The principal objective of this project was to develop an advanced, fully validated, state-of-the-science modeling system for the prediction of urban flow (i.e., turbulent flow through cities) and the concomitant problem of the modeling of the dispersion of CBRN agents released into these complex flows. This system allows the dispersion of CBRN materials to be modeled over a vast range of length scales at the appropriate resolution for each scale: namely, in the near field (up to about 2 km) where dispersion is governed by the micro-scale regime of the planetary boundary layer; to the intermediate field between about 2 and 20 km where dispersion is governed by the local or meso-γ scale; through the far field covering the range from about 20–200 km (meso-β scale) and from about 200–2000 km (meso-α scale) which correspond to dispersion at the regional scale; and, finally out to the very far field encompassing scales greater than about 2000 km corresponding to dispersion on the large (synoptic and global) scales.

The multiscale modeling system for emergency response developed in CRTI Project 02-0093RD consists of five major components shown in the schematic diagram of Figure 1. These five components can be described briefly as follows. Component 1 involves the development of models to predict the mean flow and turbulence in the urban complex at the microscale (from the building and street scale up to a length scale of about 2 km). Two kinds of models have been developed for this purpose: namely, high-resolution building-aware models for urban flow where buildings are explicitly resolved; and, virtual building models for urban flow where groups of buildings are represented simply in terms of a distributed drag force. The resulting flow solver is known as urbanSTREAM and provides predictions of the high-resolution wind and turbulence fields in an arbitrary urban environment using a Reynolds-averaged Navier-Stokes (RANS) approach.

CRTI-02-0093RD
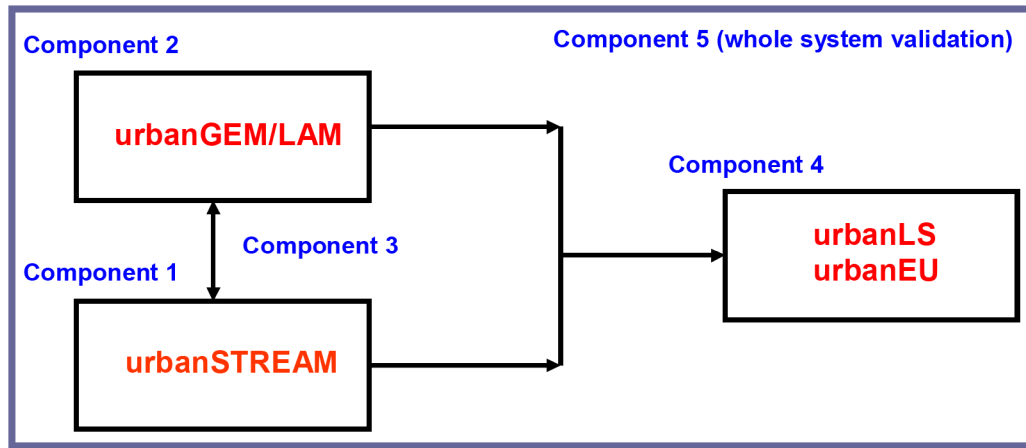
Figure 1: *Relationships between various components of CRTI Project 02-0093RD.*

Component 2 involves the inclusion of the effects of urban land use and land cover on the subgrid scales of a mesoscale meteorological model through an urban parameterization. This parameterization is required to account properly for the area-averaged effects of form drag, increased turbulence production, heating and surface energy budget modifications due to the presence of buildings/obstacles and consequences of land use and land cover within the urban environment. The resulting "urbanized" large-scale environmental flow model is known as urbanGEM/LAM (derived from Global Environmental Multiscale or GEM model, and its limited area version GEM/LAM model). Component 3 involves coupling the urban microscale flow models developed in Component 1 with the "urbanized" large-scale environmental flow model developed in Component 2. The interface between the urban microscale flow model urbanSTREAM and the "urbanized" GEM/LAM model is demanding in that the information transfer between the two models must honor physical conservation laws, mutually satisfy mathematical boundary conditions, and preserve numerical accuracy, even though the corresponding meshes might differ in structure, resolution, and discretization methodology.

Component 4 involves using the mean flow and turbulence predicted by the multiscale flow model developed in Component 3 to "drive" either an Eulerian or a Lagrangian stochastic (LS) source-oriented model for the prediction of urban dispersion of CBRN agents. The source-oriented Eulerian model for urban dispersion, known as urbanEU, is based on the numerical solution of a K-theory advection-diffusion equation. The source-oriented LS model for urban dispersion, known as urbanLS, computes the forward trajectories of "marked" fluid parcels released from a transient or continuous source and are "driven" using the full three-dimensional building-resolving wind field provided by urbanSTREAM. Finally, Component 5 involves a validation of the multiscale modeling system for both the urban flow and dispersion components. For a more detailed technical description of the various components developed under CRTI Project 02-0093RD, the reader is referred to references

3–6.

The next major step is to transition the integrative multiscale urban modeling system described briefly above towards the status of an operational system that is fully functional within a government operations facility. To achieve this objective, CRTI Project 07-0196TD entitled "Towards an Operational Urban Modeling System for CBRN Emergency Response and Preparedness", involving a collaborative model development effort by Defence R&D Canada – Suffield and Environment Canada, was approved in 2007. The primary objective of this project is to transition the state-of-the-science urban flow/dispersion modeling system, developed under CRTI project 02-0093RD, towards the status of a functional operational system at Environment Canada's Environmental Emergency Response Section (EC-EERS).

The proposed system will provide EC-EERS with the key-enabling technology to demonstrate its capability as a primary national reach-back and support centre for CBRN preplanning, real-time emergency response, and post-incident assessment in Canada. The successful completion of CRTI Project 07-0196TD will provide EC-EERS with the science and technology (S&T) that will allow them to function as the primary Federal source of expert quality-assured CBRN dispersion predictions and concomitant decision-support aids that can be used by federal, provincial and municipal agencies and emergency planners and first responders (civilian and military) for informed response decision making for mission support in both domestic and international operations, as well as for support to major events of national and international significance (e.g., Vancouver Winter Olympics in 2010). Furthermore, this development is in direct alignment with Defence R&D Canada's S&T Functional Planning Guidance in the defence and security domain to "build a reusable major events security capability" and "to provide a rigorous foundation for national defence and security emerging concepts and doctrine".

To achieve the objectives of CRTI Project 07-0196TD, the effort will be directed to three major areas: namely, (1) advanced modeling capability; (2) infrastructure for the development of supporting data and tools required for the operational system; and, (3) demonstration of the operational system. The first area of the project involves providing significant improvements to the multiscale urban modeling system exhibited in Figure 1. This effort is focused specifically on the following areas: improvements to the urban parameterization scheme in the urbanGEM/LAM model; incorporation of thermal effects in the building-resolving urbanSTREAM model; and, development of techniques for the fusion of CBRN sensor data with model predictions for source reconstruction. The second area concerns the acquisition and development of the datasets and functionalities required to transition the urban modeling system to a fully operational demonstration status. This requires provision of supporting databases such as urban building and morphology data over all major Canadian cities, databases of population distribution, hazard material source characteristics for various CBRN release modes (e.g., improvised dispersion devices, sprayers, etc.), and CBRN material and toxicological properties. Finally, the third area of the project involves demonstrating and exercising the operational system for a number of CBRN scenarios in various major Canadian cities, including participation in various national-to-local exercises (e.g., support for Vancouver Winter Olympics) and improving the modeling system products

through user feedback from the first-responder community.

A series of reports and user's guides describing the work conducted on the three major areas of CRTI Project 07-0196TD will be prepared as the effort progresses. The primary objective of this report is to describe the present status (and, more specifically, the technical formulation) of the ongoing model development conducted under one aspect of the advanced modeling capability area of the project: namely, the development of an innovative sensor-driven modeling paradigm for source reconstruction (viz., the determination of the characteristics of an unknown source following event detection by a network of CBRN detectors).

1.2 Motivation

An increasingly capable sensing technology for concentration measurements of contaminants (such as CBRN agents) released into the turbulent atmosphere, either accidentally or deliberately, has fostered interest in exploiting this information for detection, identification and reconstruction of pollutant (contaminant) sources responsible for the observed concentration. A principal impediment and critical capability gap in current emergency management efforts, directed at terrorist incidents involving a covert (clandestine) release of a CBRN agent in a densely populated urban centre, is the determination of the unknown source characteristics following event detection by a (usually limited) network of CBRN detectors. A sensor array may indicate that a putative release has occurred, but without knowledge of the unknown source characteristics of the release necessary to perform a transport and dispersion calculation, the event detection provides no more useful information than a failure indicator light. The fusing of sensor data with atmospheric transport and dispersion modeling will lead to a greatly improved situational awareness in the battlespace and result in a significantly enhanced common operating picture required for informed CBRN response decision making.

This perspective underpins the deployment by the Department of Homeland Security of (albeit sparse) arrays of biological agent sensors in 31 cities across the United States in order to provide detection, warning and reporting of a covert bioterrorism event (with plans to expand to 120 as part of the BioWatch program [7]). The BioWatch program has provided the impetus for recent research efforts directed towards a solution of the source reconstruction problem for inferring the location and emission rate of the source of contamination associated with a clandestine release of a possible biological warfare agent. Certainly, determination of the characteristics of the unknown source is perhaps the most critical information required by emergency responders for the delineation of hazard zones (toxic corridors) resulting from the contaminant release and for implementation of an appropriate mitigation strategy (e.g., identification of exposed individuals, formulation of decisions for prophylatic treatment in the case of biological agents) required to counter the CBRN agent release. Further motivation is provided by a network of 40 radiological detectors that has been set up as a verification tool for the Comprehensive Test Ban Treaty (CTBT) in order to provide world-wide monitoring of radioactive noble gases that could be used potentially for source localization and characterization of a clandestine nuclear test [8, 9].

CRTI-02-0093RD
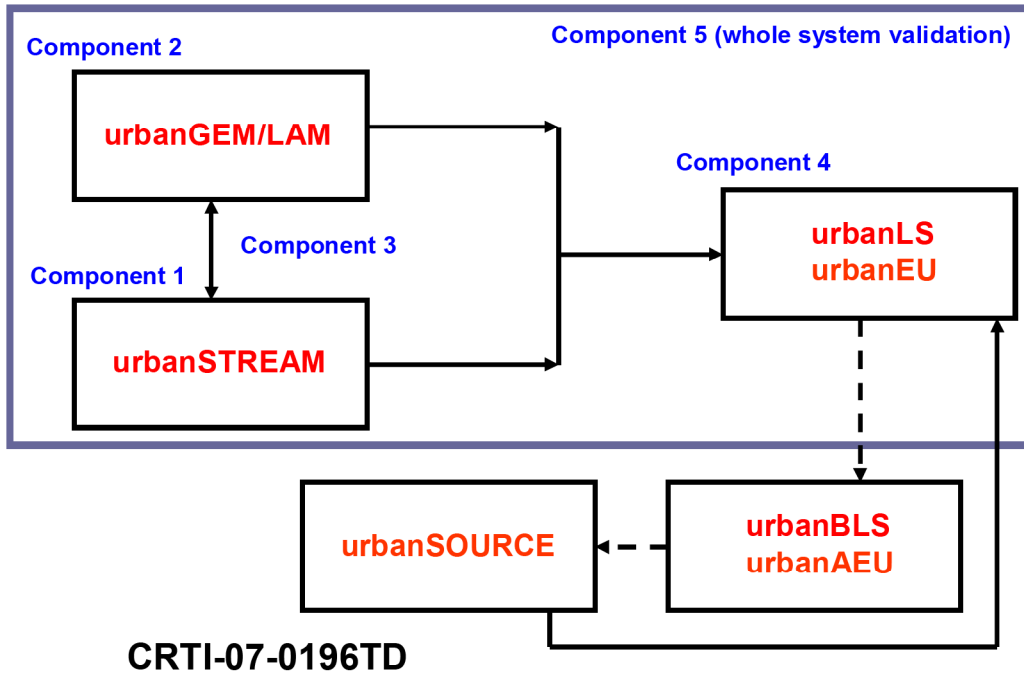
CRTI-07-0196TD

Figure 2: *Relationships of the urbanSOURCE module to the various components developed in CRTI Project 02-0093RD.*

In view of this, one major objective of CRTI Project 07-0196TD is to provide robust software that integrates CBRN sensor measurements with atmospheric (and more particularly urban) dispersion models to determine the initial source characteristics, which would allow a subsequent timely prediction of the agent dispersal in the atmospheric (urban) environment to enable informed decisions to be made on how best to respond to and mitigate the consequences of the putative agent release. To this end, we have developed a software module named urbanSOURCE which implements a Bayesian probabilistic inferential framework for source reconstruction. The relationship of this module to the various components developed in CRTI-02-0093RD is summarized in Figure 2. The source-receptor relationship required in urbanSOURCE can in principle be obtained by using either the source-oriented Eulerian or LS dispersion models urbanEU or urbanLS, respectively. However, for the Bayesian inversion of concentration data to be practical, fast and efficient techniques are required for the determination of the source-receptor relationship when the sensor concentration at a fixed receptor is of principal interest for a range of different emission scenarios (as is the case for source reconstruction). In consequence, we have developed specialized versions of urbanEU and urbanLS for use with urbanSOURCE; namely, we have implemented a receptor-oriented Eulerian dispersion model urbanAEU and a receptor-oriented LS dispersion model urbanBLS. These models are simply the adjoint representations of the source-receptor relationships embodied in the source-oriented models urbanEU and urbanLS, respectively (with more details provided in Section 2).

The science and technology that underpins the implementation of urbanSOURCE is based on an innovative Bayesian inferential methodology that is used in conjunction with the adjoint method for representation of the source-receptor relationship. Over the past three years, this methodology for source reconstruction has been formulated, developed and refined by Yee [10–15], Keats [10], Yee et al. [17, 18] and Keats et al. [19, 20] for the determination of unknown parameters of single or multiple sources for dispersion of conservative and non-conservative scalars in simple (level, unobstructed terrain) and complex (e.g., urban terrain, complex terrain on continental scales) environments. The Bayesian framework provides the proper method to deal with incomplete and noisy concentration data in the source reconstruction problem, and furthermore permits the rigorous determination of the uncertainty in the inference of the source parameters, hence extending the potential of the methodology as a tool for quantitative source reconstruction.

2 Dispersion modeling for source reconstruction

To apply Bayesian probability theory to source reconstruction, we need to relate the hypotheses of interest about the (unknown) source distribution (e.g., the source distribution is localized at a specific location and is continuously releasing material at a given emission rate) to the available concentration data measured by the array of sensors. This requires the calculation of a modeled (predicted) mean concentration $\overline{C}$. Towards this objective, we need to specify a source-receptor relationship (or, atmospheric dispersion model) that encodes how the source parameter hypotheses are related to the concentration data. More specifically, the source-receptor relationship is a mapping $\mathcal{M}_{\mathrm{SR}}$ from the hypothesis space $\mathcal{H}$ of source distributions to the sample space $\mathcal{S}^N$ of concentration data so $\mathcal{M}_{\mathrm{SR}} : \mathcal{H} \to \mathcal{S}^N$, where N is the number of concentration data.

Let the concentration at a spatial location $\mathbf{x} \equiv (x, y, z)$ and at time t be denoted $C(\mathbf{x}, t)$. The mean concentration "seen" by a sensor corresponds to an average of $C(\mathbf{x},t)$ over the sensor volume and sampling time and is given by

$$\overline{C}(\mathbf{x}_r, t_r) \equiv \int_0^T dt \int_{\mathcal{D}} d\mathbf{x} C(\mathbf{x}, t) h(\mathbf{x}, t | \mathbf{x}_r, t_r) \equiv \langle C, h \rangle (\mathbf{x}_r, t_r), \qquad (1)$$

where $h(\mathbf{x}, t | \mathbf{x}_r, t_r)$ is the spatial-temporal filtering function for the sensor located at $\mathbf{x}_r$ at time t_r, and $\mathcal{D} \times [0, T]$ corresponds to a space-time volume that contains the source and the receptors (sensors). The spatial-temporal filtering function h is constrained as follows:

$$\int_0^T dt \int_{\mathcal{D}} d\mathbf{x} h(\mathbf{x}, t | \mathbf{x}_r, t_r) = 1. \qquad (2)$$

Note that in Eq. (1), we can express the mean concentration $\overline{C}(\mathbf{x}_r, t_r)$ "seen" by a sensor as the inner (or scalar) product $\langle C, h \rangle$ of the mean concentration C and sensor response function h.

In principle, a *source-oriented* approach can be used to characterize the mapping noted above. For example, the source-oriented Eulerian dispersion model urbanEU solves an

advection-diffusion equation of the form

$$\frac{\partial C}{\partial t} + \mathbf{U} \cdot \nabla C - \nabla \cdot (K \nabla C) = S \tag{3}$$

for $C(\mathbf{x},t)$ forward in time for a given contaminant source density function $S \equiv S(\mathbf{x}, t)$, following which Eq. (1) can be used to determine the modeled mean concentration seen by a sensor at $\mathbf{x}_r$ at time t_r as the inner product $\langle C, h \rangle$. In Eq. (3), $\mathbf{U} \equiv \mathbf{U}(\mathbf{x}, t)$ is the (Eulerian) mean velocity field in the flow domain and K is a turbulent diffusivity used to model the turbulent scalar fluxes.

Alternatively, the source-oriented LS dispersion model urbanLS can be used to obtain $C(\mathbf{x},t)$. In this approach, C is estimated from the statistical characteristics of particle trajectories modeled using the following stochastic differential equation [21]:

$$
\begin{aligned}
d\mathbf{X}(t) &= \mathbf{V}(t)\, dt, \\
d\mathbf{V}(t) &= \mathbf{a}(\mathbf{X}(t), \mathbf{V}(t), t)\, dt + \left(C_0 \epsilon(\mathbf{X}(t), t)\right)^{1/2} d\mathbf{W}(t),
\end{aligned}
\tag{4}
$$

where $\mathbf{X}(t) = (X_i(t)) = (X_1(t), X_2(t), X_3(t))$ and $\mathbf{V}(t) = (V_i(t)) = (V_1(t), V_2(t), V_3(t))$ are the (Lagrangian) position and velocity, respectively, of a "marked" fluid element (or, particle) at time t (marked by the source as the fluid element passes through it at some earlier time t'), so $(\mathbf{X}, \mathbf{V})$ determines the state of the fluid particle at any time t after its initial release from the source distribution S. In Eq. (4), C_0 is the Kolmogorov "universal" constant (associated with the Kolmogorov similarity hypothesis for the form of the second-order Lagrangian velocity structure function in the inertial subrange); ϵ is the mean dissipation rate of turbulence kinetic energy; $d\mathbf{W}(t) \equiv (dW_i(t)) - (dW_1(t), dW_2(t), dW_3(t))$ are the increments of a vector-valued (three-dimensional) Wiener process; and $\mathbf{a} \equiv (a_i) = (a_1, a_2, a_3)$ is the drift coefficient vector (or, more precisely, the conditional mean acceleration vector).

Unfortunately, the source-oriented approach is computationally expensive for use in a Bayesian inferential procedure for source reconstruction, since sampling from the posterior distribution of the source parameters will potentially involve consideration of a large number of source parameter hypotheses. Each one of these hypotheses requires the solution either of an advection-diffusion equation for the Eulerian description, or of a stochastic differential equation for the Lagrangian description of atmospheric diffusion. In other words, each new release scenario defined by a source parameter hypothesis will define a different source density function S, with the consequence that either Eqs. (3) or (4) will need to be solved once for each source parameter hypothesis. This is highly computer intensive, as the simulation-based Bayesian inference procedure requires a large number of forward calculations of the mean concentration to be performed, each of which can potentially take several minutes on a modern personal computer.

In view of this, it is more computationally efficient to apply a *receptor-oriented* approach for representation of the source-receptor relationship for use in the Bayesian inference scheme. To this end, the modeled mean concentration $\overline{C}(\mathbf{x}_r, t_r)$ can be computed using the following

dual system representation:

$$
\begin{aligned}
\overline{C}(\mathbf{x}_r, t_r) &\equiv \int_0^T dt \int_{\mathcal{D}} d\mathbf{x} \, C(\mathbf{x}, t) h(\mathbf{x}, t | \mathbf{x}_r, t_r) \\
&= \int_0^{t_r} dt_0 \int_{\mathcal{D}} d\mathbf{x}_0 \, C^*(\mathbf{x}_0, t_0 | \mathbf{x}_r, t_r) S(\mathbf{x}_0, t_0) \equiv \langle C^*, S \rangle (\mathbf{x}_r, t_r),
\end{aligned} \tag{5}
$$

where $C^*(\mathbf{x}_0, t_0 | \mathbf{x}_r, t_r)$ is the adjunct (or dual) concentration at space-time point $(\mathbf{x}_0, t_0)$ associated with the concentration data at location $\mathbf{x}_r$ at time t_r (with $t_0 \leq t_r$). A comparison of Eqs. (1) and (5) implies the duality relationship $\langle C, h \rangle = \langle C^*, S \rangle$ between C and C^* through the source functions h and S. Moreover, C^* is uniquely defined in the sense that it is explicitly constructed so that it verifies this duality relationship.

In general, the adjunct concentration C^* can be obtained in the Eulerian description from the solution of the adjoint of an advection-diffusion equation with the "source" term h (space-time filtering imposed by the sensor) as follows:

$$
-\frac{\partial C^*}{\partial t} - \mathbf{U} \cdot \nabla C^* - \nabla \cdot (K \nabla C^*) = h. \tag{6}
$$

Equation (6) is the key result underpinning the receptor-oriented Eulerian dispersion model urbanAEU (see Figure 2). In the receptor-oriented approach, Eq. (6) is solved backwards in time for a given receptor, and the concentration 'seen' by the sensor at this receptor can be calculated directly using Eq. (5) once the source distribution S has been specified [3]. It should be noted that the computations can be repeated for any source distribution S to obtain the concentration at the given receptor without having to re-solve Eq. (6) for C^*. In consequence, the receptor-oriented approach is ideally suited for source reconstruction using Bayesian inference.

Equivalently, in the Lagrangian description, the adjunct field C^* can be determined using a backward-time Lagrangian trajectory model, defined as the solution to the following stochastic differential equation (with $dt' > 0$):

$$
\begin{aligned}
d\mathbf{X}^b(t') &= \mathbf{V}^b(t') \, dt', \\
d\mathbf{V}^b(t') &= \mathbf{a}^b(\mathbf{X}^b(t'), \mathbf{V}^b(t'), t') \, dt' + \left(C_0 \epsilon (\mathbf{X}^b(t'), t') \right)^{1/2} d\mathbf{W}(t'),
\end{aligned} \tag{7}
$$

where at any given time t', $(\mathbf{X}^b, \mathbf{V}^b)$ is a point in the phase space along the backward trajectory of the "marked" fluid element (here assumed to be marked or tagged as a fluid particle which at time t_r passed through the spatial volume of the detector at location $\mathbf{x}_r$). This result is the fundamental equation underpinning the receptor-oriented LS model urbanBLS (see Figure 2). In this receptor-oriented approach, "marked" fluid elements with final space-time coordinates $(\mathbf{x}_f, t_f)$ are sampled from a space-time density function that is proportional to the (prescribed) spatial-temporal filtering function $h(\mathbf{x}', t' | \mathbf{x}_r, t_r)$ at a receptor space-time location $(\mathbf{x}_r, t_r)$. The backward-time Lagrangian trajectories $(t' < t_f)$ of these "marked" fluid elements, which emanate from the receptor space-time volume and move towards the source (so, $t' \to t' - dt'$ with $dt' > 0$), are determined in accordance to

Eq. (7) and the displacement statistics of these "marked" fluid elements can be used to compute $C^*(\mathbf{x}', t'|\mathbf{x}_r, t_r)$ (which is interpreted here as a function of $\mathbf{x}'$ and t').

It can be shown ([21], [22]) that C^* obtained from Eq. (7) for a detector with the filtering function h and C obtained from Eq. (1) for a release from the source density S, are exactly consistent with the duality relationship $\langle C, h \rangle = \langle C^*, S \rangle$ provided: (1) $\mathbf{V}^b$ in Eq. (7) is related to $\mathbf{V}$ in Eq. (4) as $\mathbf{V}^b(t) = \mathbf{V}(t)$; and, (2) $\mathbf{a}^b$ in Eq. (7) is related to $\mathbf{a}$ in Eq. (4) as

$$a_i^b(\mathbf{x}, \mathbf{u}, t) = a_i(\mathbf{x}, \mathbf{u}, t) - C_0 \epsilon(\mathbf{x}, t) \frac{\partial}{\partial u_i} \ln P_E(\mathbf{u}), \qquad (8)$$

where $P_E(\mathbf{u})$ is the background Eulerian velocity PDF, which in urbanBLS is assumed to have a Gaussian form [5].

3   Model for concentration observations

The models described above provide predictions for the "ideal" mean concentration seen by a sensor at the receptor space-time point $(\mathbf{x}_r, t_r)$. The actual concentration data measured by the sensor will not usually agree with the concentration predicted by the model owing to the "noise" process imposed on both the measured and predicted concentration data, which by its very nature is expected to possess a very complicated structure. To this purpose, it is assumed that the actual concentration data available from the network of sensors are measured at a finite number of sensor locations and at a finite number of time points at each sensor location. The actual concentration datum $d_{i,j^{(i)}}$ acquired by the sensor at receptor location $\mathbf{x}_{r_i}$ and at time $t_{r_j}^{(i)}$ ($i = 1, 2, \ldots, N_d$ and $j = 1, 2, \ldots, N_t^{(i)}$, where N_d is the number of sensors and $N_t^{(i)}$ is the number of time samples measured at the i-th sensor) is assumed to be the sum of a modeled mean concentration signal $\overline{C}(\mathbf{x}_{r_i}, t_{r_j}^{(i)}; \Theta)$ and "noise" $e_{i,j^{(i)}}$, so

$$d_{i,j^{(i)}} = \overline{C}(\mathbf{x}_{r_i}, t_{r_j}^{(i)}; \Theta) + e_{i,j^{(i)}}, \qquad (9)$$

where Θ is an appropriate parameter vector describing the source distribution S; and, $\overline{C}(\mathbf{x}_r, t_r; \Theta)$ is the modeled mean concentration at location $\mathbf{x}_r$ and time t_r, determined in accordance to Eq. (5) for a source distribution characterized by parameter vector Θ. For simplicity of notation, the variables in Eq. (9) which are indexed or labelled by $(i, j^{(i)})$ will be ordered in some regular and convenient manner (e.g., lexicographic ordering) and this collection will be indexed by J ($J = 1, 2, \ldots, N$, with $N \equiv \sum_{i=1}^{N_d} N_t^{(i)}$ being the total number of measured concentration data). Then, we can write the observation model as follows:

$$d_J = \overline{C}_J(\Theta) + e_J, \qquad J = 1, 2, \ldots, N, \qquad (10)$$

where $\overline{C}_J(\Theta) \equiv \overline{C}(\mathbf{x}_{r_i}, t_{r_j}^{(i)}; \Theta)$. In Eq. (10), e_J is a noise term representing the discrepancy between the measured concentration d_J and the predicted concentration $\overline{C}_J(\Theta)$.

In general, e_J consists of errors (e.g, input, stochastic, and measurement) and any real signal in the data that cannot be explained by the model. The random error e_J can be split

into four terms as discussed by Rao [23], so

$$e_J = \eta_J^{(1)} + \eta_J^{(2)} + \eta_J^{(3)} + \eta_J^{(4)}. \tag{11}$$

The first term $\eta_J^{(1)}$ of the error corresponds to model error arising from uncertainties in the representation of various physical processes in the dispersion model used to predict the mean concentration. The second term $\eta_J^{(2)}$ describes the input error arising from uncertainties in the values of empirical parameters and/or specification of the input meteorology (initial and boundary conditions) used by the dispersion model. The third term $\eta_J^{(3)}$ of the error is the stochastic uncertainty arising from the turbulent nature of the atmosphere, which gives rise naturally to random concentration fluctuations in hazardous gas releases ([24],[25],[26]). Finally, $\eta_J^{(4)}$ describes the noise inherent in the sensor (essentially measurement or instrument error). Note that $\eta^{(1)}$ and $\eta^{(2)}$ are sources of error that affect the predicted (model) concentration, whereas $\eta^{(3)}$ and $\eta^{(4)}$ are sources of error that affect the measured concentration. All four sources of error contribute to the expected discrepancy between the measured and predicted concentration.

Rao [23] discusses the nature of these four types of error with respect to characterization of uncertainties in atmospheric dispersion models, and provides a comprehensive review of sensitivity and/or uncertainty analysis methods that have been used to quantify and reduce them. In this paper, all the various error contributions to the noise term are simply lumped together and denoted by e_J [see Eqs. (10) and (11)]. It is assumed that the observer does not have a detailed knowledge of the probability distribution of the noise (aggregate error), other than that the observer has an estimate (perhaps crude) for the expected scale of variation of the noise. More specifically, it is assumed that the noise scale parameter associated with e_J (aggregate error of the J-th concentration observation) is provided in the form of a (finite) variance σ_J^2.

To complete the specification of the source-receptor relationship, we need to consider a functional form for the source density function S. In this paper, we consider a transient point source located at $\mathbf{x}_s$ with source on and off times T_b and T_e, respectively, between which the source is releasing contaminant at a constant emission rate Q, so

$$S(\mathbf{x}, t) = Q\delta(\mathbf{x} - \mathbf{x}_s)\Big[U(t - T_b) - U(t - T_e)\Big], \tag{12}$$

where $\delta(\cdot)$ and $U(\cdot)$ are the Dirac delta and Heaviside unit step functions, respectively. It is convenient to define $\Theta \equiv (\mathbf{x}_s, T_b, T_e, Q)$ as the collection of source parameters for the source distribution given by Eq. (12). This source model has a number of important special cases. For example, defining $\Delta t \equiv T_e - T_b$ and taking the limit as $\Delta t \to 0$ in Eq. (12), then $T_e \to T_b \equiv T_s$ and we recover an instantaneous point source released at $\mathbf{x}_s$ at time T_s. Also, for a continuous point source, $T_b \to -\infty$ and $T_e \to \infty$ and the only relevant parameters for this case are $\mathbf{x}_s$ and Q.

With the formulation above, the problem of source reconstruction reduces to the following: given the observed vector of concentration data $\mathbf{D} \equiv (d_1, d_2, \ldots, d_N)$, the objective

is to estimate Θ. The problem of estimating Θ will be addressed using Bayesian probability theory which defines probability not as a frequency of occurrence but rather, as a reasonable degree of belief. The underlying basis of Bayesian probability theory provides a rigorous mathematical framework for making inferences about the source parameters and, as a consequence, provides a rigorous basis for quantifying the uncertainties in the estimated source parameters. The foundations of probability theory when interpreted as a quantitative theory of inference is summarized in the next section.

4   Probability theory as logic

The assessment of uncertainty, which is of fundamental importance to quantitative science, can be dealt with rigorously using probability calculus. The rules for this calculus can be derived fully starting with the formulation of a small number of desiderata for a theory of plausibility as first provided by Cox [27], with an eloquent description of the complete development given by Jaynes [28] in his definitive treatise.[1] This theory can be interpreted as *the* extension of Aristotelian deductive logic to cases where there is uncertainty, and is based on three basic desiderata: namely, (1) degrees of plausibility are represented by real numbers; (2) qualitative consistency with common sense so that the resulting theory will reduce properly to the rules for Aristotelian deductive logic in cases where the propositions are either certainly true or certainly false; and, (3) internal consistency in the sense that two different methods of calculation permitted by the theory give the same result.

These desiderata imply two "axioms" for the theory of plausibility (or, probability); namely, the sum rule

$$P(H|I) + P(\overline{H}|I) = 1, \tag{13}$$

and the product rule

$$P(H, D|I) = P(H|I)P(D|H, I) = P(D|I)P(H|D, I), \tag{14}$$

where $P(\cdot)$ denotes the real number measure of the plausibility (probability) of a proposition or hypothesis. These are simply the ordinary rules of probability calculus and imply every allowed (consistent) plausibility theory must be mathematically equivalent to probability theory, or else inconsistent (no other calculus is admissible for inference with the above mentioned desiderata). In probability theory as logic, a probability is reinterpreted to represent a state of knowledge, rather than a state of nature. Equation (14) can be rearranged immediately to give Bayes' theorem:

$$P(H|D, I) = \frac{P(H|I)P(D|H, I)}{P(D|I)}. \tag{15}$$

In Eqs. (13) and (14), H, D, and I denote propositions which for our application have the following explicit meanings: I denotes the background (contextual) information that defines

---

[1]Interestingly, Jaynes [28] has remarked that the article by Cox [27] is "the most important advance in the conceptual (as opposed to purely mathematical) formulation of probability theory since Laplace".

the source reconstruction problem; D denotes the concentration data made available by the array of sensors; and, H denotes a hypothesis about the (unknown) source distribution (e.g., the location of a source lies in a particular region of space and its emission rate assumes values in a particular range). The probability $P(H|I)$ is a measure of the degree to which proposition H is supported by the information embodied in proposition I, with "$|$" denoting "conditional upon". Finally, $\overline{H}$ and H, D denote "not H" and "H and D", respectively.

With this interpretation of H, D and I, the terms in Bayes' theorem of Eq. (15) are as follows. Firstly, $P(H|I)$ is the *prior* probability for a hypothesis H about the source predicated on the contextual information specified by I and encodes all the prior information about the source before receipt of the concentration data D. Secondly, $P(D|H, I)$ is the *likelihood* function and is the probability that we observe the concentration data D when H is known. Thirdly, $P(D|I)$ is referred to as the *evidence* and in our case here (which deals with parameter estimation) is simply a normalization constant and need not be considered further. Finally, $P(H|D, I)$ is the *posterior* probability for the hypothesis H about the source in light of the new information introduced through the newly acquired concentration data D.

5 Prior, likelihood, and posterior

Given the background in Section 4, the goal is to compute the joint posterior probability density function (PDF) for the source parameters Θ, which in accordance to Eq. (15) can be recast in the following form:

$$p(\Theta|\mathbf{D}, I) = \frac{p(\Theta|I)p(\mathbf{D}|\Theta, I)}{p(\mathbf{D}|I)}, \tag{16}$$

where we have made the following identifications: (1) $P(H|I) \equiv P(\Theta|I) = p(\Theta|I)\, d\Theta$; (2) $P(D|H, I) \equiv P(\mathbf{D}|\Theta, I) = p(\mathbf{D}|\Theta, I)\, d\mathbf{D}$; (3) $P(D|I) \equiv P(\mathbf{D}|I) = p(\mathbf{D}|I)\, d\mathbf{D}$; and, (4) $P(H|D, I) \equiv P(\Theta|\mathbf{D}, I) = p(\Theta|\mathbf{D}, I)\, d\Theta$. It is implicitly assumed that all probability measure admit a probability density function (so, for example, the prior can be represented either through the probability measure $P(\Theta|I)$ or its associated probability density function $p(\Theta|I)$, etc.). All the terms in Eq. (16) are to be interpreted given the background information I (e.g., background meteorology, source-receptor relationship, etc.).

To compute the joint posterior PDF of the parameters $p(\Theta|\mathbf{D}, I)$, we need to evaluate three terms in Eq. (16); namely, the prior PDF $p(\Theta|I)$, the likelihood function $p(\mathbf{D}|\Theta, I)$, and the evidence (also, frequently referred to as the prior predictive or marginal likelihood) $p(\mathbf{D}|I)$. It should be noted that for the parameter estimation problem considered in this report, the evidence $p(\mathbf{D}|I)$ is independent of Θ and simply plays the role of a normalization constant, which can be ignored. In consequence, if the posterior PDF is normalized at the end of the calculation, we get simply

$$p(\Theta|\mathbf{D}, I) \propto p(\Theta|I)p(\mathbf{D}|\Theta, I). \tag{17}$$

The specification of the posterior PDF for the source parameters Θ now reduces to the assignment of the prior PDF $p(\Theta|I)$ and the likelihood function $p(\mathbf{D}|\Theta, I)$.

5.1 Assignment of prior probability

Let us consider the assignment of the prior PDF $p(\Theta|I)$. To this purpose, we assume the logical independence of the various source parameters. This assumption implies simply that knowing the x_s location of the source tells us nothing about the y_s and z_s location of the source, knowing the location of the source tells us nothing about the source's emission rate or when it was turned on/off, etc. In other words, there is no physical reason why the various source parameters $\mathbf{x}_s$, T_b, T_e, and Q are correlated. As a consequence, the prior PDF $p(\Theta|I)$ may be factored (by repeated application of the product rule of probability calculus) to obtain

$$p(\Theta|I) = p(\mathbf{x}_s|I)p(Q|I)p(T_b|I)p(T_e|T_b, I). \tag{18}$$

In words, the joint prior PDF of the source parameters is the product of the prior PDFs for the individual parameters.

Each of the component prior PDFs in Eq. (18) for the individual source parameters can be assigned by stating what is known about them. The prior PDF for the source location $\mathbf{x}_s$ will be assigned a uniform distribution over some large region $\mathcal{D} \subset \mathbb{R}^3$:

$$p(\mathbf{x}_s|I) = \mathbb{I}_{\mathcal{D}}(\mathbf{x}_s)/\mathcal{V}(\mathcal{D}), \tag{19}$$

where $\mathcal{V}(\mathcal{D})$ is the volume of the region $\mathcal{D}$ and $\mathbb{I}_A(x)$ denotes the indicator function for a set A (viz., $\mathbb{I}_A(x) = 1$ if $x \in A$ and $\mathbb{I}_A(x) = 0$ if $x \notin A$).

It is assumed that the emission rate Q has an *a priori* upper bound $Q_{\max}$. No additional prior information on Q is assumed and a uniform prior PDF is assigned to the emission rate:

$$p(Q|I) = \mathbb{I}_{(0,Q_{\max})}(Q)/Q_{\max}. \tag{20}$$

Finally, the prior PDFs for the source on T_b and source off T_e times must be assigned. To this purpose, the prior distributions for T_b and T_e are assigned uniform distributions with the following forms:

$$p(T_b|I) = \mathbb{I}_{(0,T_{\max})}(T_b)/T_{\max}, \tag{21}$$

and

$$p(T_e|T_b, I) = \mathbb{I}_{(T_b,T_{\max})}(T_e)/(T_{\max} - T_b). \tag{22}$$

Here, $T_{\max}$ is an upper bound on the time at which the source was turned on or off. Different upper bounds can be chosen for T_b and T_e in the prior PDFs of Eqs. (21) and (22), but for the formulation in this report we simply used a common upper bound for the source on and off times (with effectively no loss in generality). Note that the time that the source is turned off must necessarily occur after it has been turned on, and this information is encoded in the form of the prior PDF for T_e given by Eq. (22), where the distribution is seen to be conditioned on T_b.

With these assignments for the component prior PDFs, the joint prior PDF for Θ given by Eq. (18) assumes the following explicit form:

$$p(\Theta|I) = \frac{\mathbb{I}_{\mathcal{D}}(\mathbf{x}_s)}{\mathcal{V}(\mathcal{D})} \cdot \frac{\mathbb{I}_{(0,Q_{\max})}(Q)}{Q_{\max}} \cdot \frac{\mathbb{I}_{(0,T_{\max})}(T_b)}{T_{\max}} \cdot \frac{\mathbb{I}_{(T_b,T_{\max})}(T_e)}{(T_{\max} - T_b)}. \tag{23}$$

5.2 Assignment of likelihood function

Next, let us consider the assignment of a functional form for the likelihood function $p(\mathbf{D}|\Theta, I)$. The likelihood function is equivalent to the direct probability for the concentration data $\mathbf{D}$, given the source parameters Θ. In urbanSOURCE, three different model equations for the likelihood function have been implemented.

In the absence of a detailed knowledge of the noise distribution e_J [cf. Eq. (11)], other than that it has a finite variance σ_J^2, the application of the principle of maximum entropy [28] informs us that a Gaussian distribution is the most conservative choice for the direct probability of the data $\mathbf{D}$ (or, equivalently, of the noise $\mathbf{e} \equiv (e_1, e_2, \ldots, e_N)$). The entropy of the PDF of the noise is a measure of its information content (viz., it is the asymptotic measure of the size of the basic support set of the distribution or 'volume' occupied by the sensibly probable noise values). The principle of maximum entropy is applied to ensure that the PDF representing our 'state of information' about the noise values does not encapsulate unwarranted assumptions (e.g., about higher-order moments of the noise which are not available). Choosing a distribution for the noise that provides the largest support set permitted by the information allows the largest range of possible variations in the noise values consistent with the available information (implying the most conservative estimates for these values).

From these considerations, the first model for the likelihood function has the following form [in light of Eq. (10)]:

Model 1

$$p(\mathbf{D}|\Theta, I) \equiv p(\mathbf{D}|\Theta, \varsigma, I) = \frac{1}{\prod_{J=1}^{N} \sqrt{2\pi}\sigma_J} \exp\left(-\frac{1}{2}\chi^2(\Theta)\right), \tag{24}$$

where

$$\chi^2(\Theta) \equiv \sum_{J=1}^{N} \left(\frac{d_J - \overline{C}_J(\Theta)}{\sigma_J}\right)^2. \tag{25}$$

In Eq. (24), the notation for the likelihood function was adjusted to include the standard deviation for the noise $\varsigma \equiv (\sigma_1, \sigma_2, \ldots, \sigma_N)$ in the conditioning to emphasize the fact that $\{\sigma_J\}_{J=1}^{N}$ are assumed to be known quantities.

As mentioned previously, the noise term e_J in Eq. (10) is extremely complicated, arising as such from a superposition of input, model, stochastic and measurement errors. In consequence, reliable estimates for σ_J ($J = 1, 2, \ldots, N$) are difficult to obtain in practical applications. In view of this, let us denote by s_J the quoted estimate of the standard deviation for the noise term e_J for which the true (but unknown) standard deviation is σ_J. Now, let us characterize the uncertainty in the specification of the standard deviation of e_J with an inverse gamma distribution of the following form:

$$p(\sigma_J|s_J, \alpha, \beta, I) = \frac{\alpha^{\beta}}{\Gamma(2\beta)} \left(\frac{s_J}{\sigma_J}\right)^{2\beta} \exp\left(-\alpha\frac{s_J^2}{\sigma_J^2}\right) \frac{1}{\sigma_J}, \quad J = 1, 2, \ldots, N, \tag{26}$$

where $\Gamma(x)$ denotes the gamma function and α and β are scale and shape parameters, respectively, that define the inverse gamma distribution.[2] Again, the parameters α and β have been added to the PDF of the noise uncertainty in Eq. (26) to indicate that the values for these parameters are assumed to be known.

In view of Eq. (26), the true but unknown standard deviations σ_J of e_J that appear in Eq. (24) can be treated as nuisance parameters and eliminated by using the product and sum rules of probability calculus to give

$$
\begin{aligned}
p(\mathbf{D}|\Theta, \mathbf{s}, \alpha, \beta, I) &= \int p(\mathbf{D}|\Theta, \varsigma, I) p(\varsigma|\mathbf{s}, \alpha, \beta, I) \, d\varsigma \\
&= \int p(\mathbf{D}|\Theta, \varsigma, I) \prod_{J=1}^{N} \frac{\alpha^{\beta}}{\Gamma(2\beta)} \left(\frac{s_J}{\sigma_J}\right)^{2\beta} \exp\left(-\alpha \frac{s_J^2}{\sigma_J^2}\right) \frac{1}{\sigma_J} \, d\varsigma, \quad (27)
\end{aligned}
$$

where $\mathbf{s} \equiv (s_1, s_2, \ldots, s_N)$ are the estimated standard deviations for the noise $(e_1, e_2, \ldots, e_N)$ and $d\varsigma \equiv d\sigma_1 d\sigma_2 \ldots d\sigma_N$. Now, substituting the form for $p(\mathbf{D}|\Theta, \varsigma, I)$ from Eqs. (24) and (25) into Eq. (27) and performing the integration with respect to ς (a process known as marginalization), we obtain the second model for the likelihood function:

Model 2

$$
p(\mathbf{D}|\Theta, \mathbf{s}, \alpha, \beta, I) = \prod_{J=1}^{N} \frac{2\alpha^{\beta}\beta}{\sqrt{2\pi}s_J} \frac{1}{\left[\alpha + \left(d_J - \overline{C}_J(\Theta)\right)^2 / (2s_J^2)\right]^{\beta+1/2}} . \quad (28)
$$

The second model for the likelihood function given by Eq. (28) depends explicitly on the hyperparameters α and β for which values need to be assigned. In urbanSOURCE, the values of α and β are assigned as $\alpha = \pi^{-1}$ (default value) and $\beta = 1$. The assignment $\beta = 1$ results in a very heavy-tailed distribution for $p(\sigma_J|s_J, \alpha, \beta, I)$ which allows significant deviations of the noise uncertainty from the quoted value of s_J (provided by the user). Indeed, with the choice $\beta = 1$, the variance associated with $p(\sigma_J|s_J, \alpha, \beta, I)$ in Eq. (26) becomes infinite. The heavy tail of the distribution is chosen to account for possibly significant under-estimations of the actual uncertainty (viz., the quoted uncertainty $s_J \ll \sigma_J$). This could arise from inconsistencies in the model concentration predictions owing to structural model error or to 'outliers' in the measured concentration data owing to either measurement error or perhaps distortion of the measured concentration data due to some unrecognized spurious source.

With a fixed value of $\beta = 1$, the parameter α is related to the bias in the estimation of σ_J using s_J. The choice $\alpha = 1$ implies that $\langle\sigma_J\rangle = s_J$ and would code for our belief that the estimates s_J of the standard deviations of e_J are unbiased. The specification $\alpha > 1$ codes for an expected negative bias in the user's estimates for the actual uncertainties σ_J (viz., $s_J < \langle\sigma_J\rangle$). Finally, the choice $\alpha < 1$ encodes for an expected positive bias in the user's estimates for the actual uncertainties σ_J (viz., $s_J > \langle\sigma_J\rangle$). In urbanSOURCE, the value for

---

[2]Note that the inverse gamma distribution for σ_J has mean $\langle\sigma_J\rangle = \alpha^{1/2}s_J/(2\beta-1)$ and variance $\text{Var}[\sigma_J] = \alpha s_J^2 / (2(2\beta-1)^2(\beta-1))$. Furthermore, the mode of the distribution is $(\sigma_J)_{\text{mode}} = (2\alpha/(1+2\beta))^{1/2}s_J$.

α is user selected, but if the user does not provide the program with a value for α a default value of $\alpha = \pi^{-1}$ is used. This particular choice of α results in a mode for the distribution of σ_J at $(\sigma_J)_{\text{mode}} \approx 0.46 s_J$.

The third model for the likelihood function is based on a modified lognormal distribution for the noise e_J suggested by Senocak et al. [29]:

$$
\begin{aligned}
p(\mathbf{D}|\Theta, \sigma, \nu, I) \;\; = \;\; & \prod_{J=1}^{N} \Bigg(\mathbb{I}_{\{d_J=0\}}(d_J) \exp\big(-\nu \overline{C}_J(\Theta)\big) + \mathbb{I}_{\{d_J>0\}}(d_J) \\
& \times \frac{\big(1 - \exp\big(-\nu \overline{C}_J(\Theta)\big)\big)}{\sqrt{2\pi} d_J \sigma} \exp\bigg(- \frac{\big(\log d_J - \log \overline{C}_J(\Theta)\big)^2}{2\sigma^2}\bigg) \Bigg),
\end{aligned} \quad (29)
$$

where the dependence on the hyperparameters ν and σ has been made explicit in the likelihood function. The parameter σ^2 is the standard deviation of the difference between the logarithm of the measured concentration and the logarithm of the modeled (predicted) concentration. The likelihood function requires knowing σ, which is usually not available in practical applications. In view of this, we treat σ as a nuisance parameter and eliminate it using the process of marginalization. To that purpose, we need to assign a prior PDF for σ, $p(\sigma|I)$. To achieve this objective, we recognize that the standard deviation is a scale parameter and that the completely uninformative prior PDF for a scale parameter is Jeffrey's prior [30] $p(\sigma|I) \propto \sigma^{-1}$. Strictly speaking, Jeffrey's prior is an improper prior in the sense that it is not normalizable over the domain of definition $(0, \infty)$. Nevertheless, the use of Jeffrey's prior for $p(\sigma|I)$ is harmless because the exponential cutoff in the lognormal distribution is so sharp that its use here results always in convergent integrals. In view of this, we can marginalize the nuisance parameter σ by multiplying the part of $p(\mathbf{D}|\Theta, \sigma, \nu, I)$ in Eq. (29) with $\mathbb{I}_{\{d_J>0\}}(d_J)$ by $p(\sigma|I) \propto \sigma^{-1}$ and integrating with respect to σ to get[3]

Model 3

$$
\begin{aligned}
p(\mathbf{D}|\Theta, \nu, I) \;\; \propto \;\; & \int_0^{\infty} p(\mathbf{D}|\Theta, \sigma, \nu, I) p(\sigma|I) \, d\sigma \\
\propto \;\; & \prod_{J'=1}^{N_0} \exp\big(-\nu \overline{C}_{J'}(\Theta)\big) \times \\
& \prod_{J''=1}^{N_>} \big(1 - \exp\big(-\nu \overline{C}_{J''}(\Theta)\big)\big) \times \Bigg[\frac{1}{2} \sum_{\tilde{J}=1}^{N_>} \big(\log d_{\tilde{J}} - \log \overline{C}_{\tilde{J}}(\theta)\big)^2 \Bigg]^{-\frac{N_>}{2}}, \quad (30)
\end{aligned}
$$

where N_0 and $N_>$ are the number of zero and non-zero concentration observations, respectively ($N = N_0 + N_>$); the product of terms over J' includes all concentration observations for which $d_J = 0$; and, the product of terms over J'' (as well as the sum of terms over $\tilde{J}$) includes all concentrations for which $d_J > 0$.

Finally, a choice for the hyperparameter ν in Eq. (30) needs to be made. To this purpose, we follow Senocak et al. [29] and fix ν as $\nu = \log(2)/C_{\text{th}}$, where C_{th} is the threshold

---

[3]A simple change of variables transforms the integral into the standard form for a gamma integral.

concentration for the sensor. With this choice for ν, the concentration in a plume or cloud is detected with a probability of $\exp(-\nu\overline{C}_J) = 1/2$ when the predicted concentration $\overline{C}_J$ at the sensor is equal to the threshold concentration C_{th}.

5.3 Posterior probability

The posterior PDF of the source parameters Θ, $p(\Theta|\mathbf{D}, I)$, can be determined by application of Bayes' rule of Eq. (17) with reference to the prior PDF for the Θ given by Eq. (23) and the three models for the likelihood function given by Eqs. (24), (28) and (30). This gives the following three model equations for the posterior PDF of Θ:

Model 1

$$
\begin{aligned}
p(\Theta|\mathbf{D}, I) \quad\equiv\quad & p(\Theta|\mathbf{D}, \varsigma, I) \propto p(\Theta|I)p(\mathbf{D}|\Theta, \varsigma, I) \\
\propto\quad & \frac{\mathbb{I}_{\mathcal{D}}(\mathbf{x}_s)}{\mathcal{V}(\mathcal{D})} \cdot \frac{\mathbb{I}_{(0,Q_{\max})}(Q)}{Q_{\max}} \cdot \frac{\mathbb{I}_{(0,T_{\max})}(T_b)}{T_{\max}} \cdot \frac{\mathbb{I}_{(T_b,T_{\max})}(T_e)}{(T_{\max} - T_b)} \\
& \times \frac{1}{\prod_{J=1}^{N}\sqrt{2\pi}\sigma_J} \exp\left(-\frac{1}{2}\sum_{J=1}^{N}\left(\frac{d_J - \overline{C}_J(\Theta)}{\sigma_J}\right)^2\right) ;
\end{aligned}
\tag{31}
$$

Model 2

$$
\begin{aligned}
p(\Theta|\mathbf{D}, I) \quad\equiv\quad & p(\Theta|\mathbf{D}, \mathbf{s}, \alpha, \beta, I) \propto p(\Theta|I)p(\mathbf{D}|\Theta, \mathbf{s}, \alpha, \beta, I) \\
\propto\quad & \frac{\mathbb{I}_{\mathcal{D}}(\mathbf{x}_s)}{\mathcal{V}(\mathcal{D})} \cdot \frac{\mathbb{I}_{(0,Q_{\max})}(Q)}{Q_{\max}} \cdot \frac{\mathbb{I}_{(0,T_{\max})}(T_b)}{T_{\max}} \cdot \frac{\mathbb{I}_{(T_b,T_{\max})}(T_e)}{(T_{\max} - T_b)} \\
& \times \prod_{J=1}^{N} \frac{2\alpha^\beta\beta}{\sqrt{2\pi}s_J} \frac{1}{\left[\alpha + (d_J - \overline{C}_J(\Theta))^2/(2s_J^2)\right]^{\beta+1/2}} ;
\end{aligned}
\tag{32}
$$

Model 3

$$
\begin{aligned}
p(\Theta|\mathbf{D}, I) \quad\equiv\quad & p(\Theta|\mathbf{D}, \nu, I) \propto p(\Theta|I)p(\mathbf{D}|\Theta, \nu, I) \\
\propto\quad & \frac{\mathbb{I}_{\mathcal{D}}(\mathbf{x}_s)}{\mathcal{V}(\mathcal{D})} \cdot \frac{\mathbb{I}_{(0,Q_{\max})}(Q)}{Q_{\max}} \cdot \frac{\mathbb{I}_{(0,T_{\max})}(T_b)}{T_{\max}} \cdot \frac{\mathbb{I}_{(T_b,T_{\max})}(T_e)}{(T_{\max} - T_b)} \\
& \times \prod_{J'=1}^{N_0} \exp\left(-\nu\overline{C}_{J'}(\Theta)\right) \times \\
& \prod_{J''=1}^{N_>} \left(1 - \exp\left(-\nu\overline{C}_{J''}(\Theta)\right)\right) \times \left[\frac{1}{2}\sum_{\tilde{J}=1}^{N_>}\left(\log d_{\tilde{J}} - \log \overline{C}_{\tilde{J}}(\theta)\right)^2\right]^{-\frac{N_>}{2}} .
\end{aligned}
\tag{33}
$$

In Eqs. (31), (32) and (33), the new parameters on which the models for the posterior distribution depend have been explicitly added after the vertical bar "|" (viz., each model for the posterior distribution is conditioned upon the new parameters).
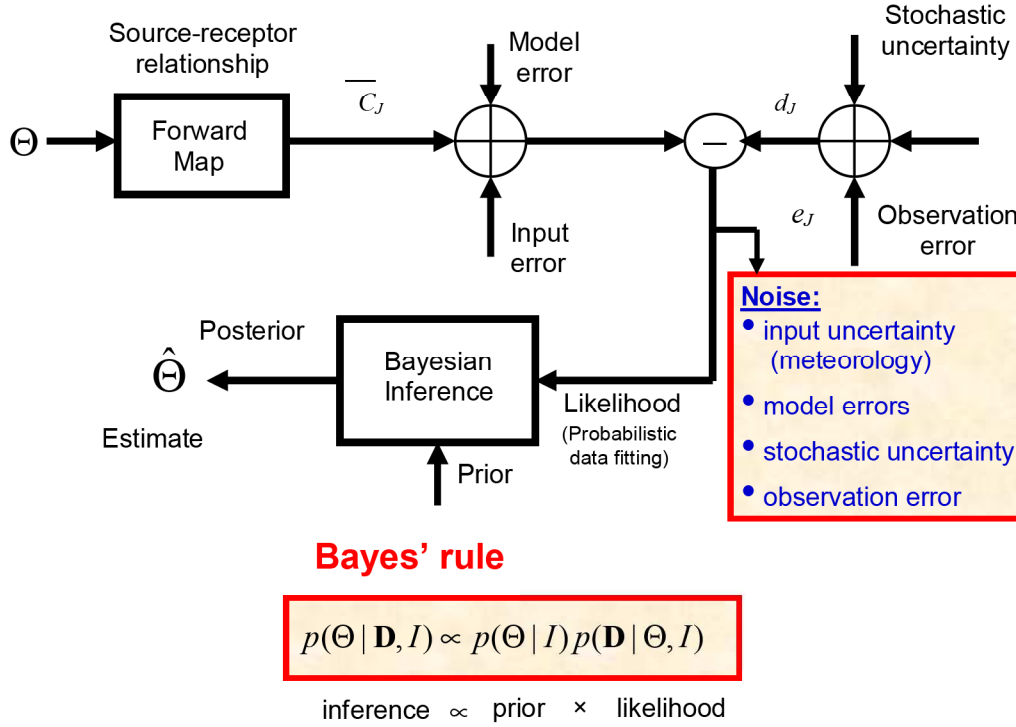
Figure 3: *Bayesian inference scheme implemented in urbanSOURCE.*

The posterior PDF $p(\Theta|\mathbf{D}, I)$ embodies the state of information about the source parameters given the prior information about these parameters encoded in the prior PDF $p(\Theta|I)$ and the newly acquired concentration data $\mathbf{D}$, which modulates our prior belief about Θ through the likelihood function $p(\mathbf{D}|\Theta, I)$. The numerical values of Θ indexes a continuous sequence of hypotheses about the unknown source distribution, and the posterior probability $p(\Theta|\mathbf{D}, I)d\Theta$ ranks the plausibility of these hypotheses. Low values of the posterior probability indicate that the numerical value of Θ is improbable (viz., the particular source distribution described by Θ is implausible), whereas high values indicate high plausibility for the hypothesis. Furthermore, $p(\Theta|\mathbf{D}, I)$ allows us to estimate all the interesting statistics about the source parameters Θ (more about this in the next section).

To summarize, urbanSOURCE computes the three models for the posterior PDF $p(\Theta|\mathbf{D}, I)$ given in Eqs. (31), (32) and (33). The components that make up urbanSOURCE are depicted in Figure 3. In particular, the predicted (model) concentration $\overline{C}_J$ $(J = 1, 2, \ldots, N)$ needed to determine the likelihood function are computed using Eq. (5), in which the dual (adjoint) concentration fields C^* are determined using either urbanAEU (for an Eulerian description of the dispersion) or urbanBLS (for a Lagrangian description of the dispersion).

The relationship between urbanSOURCE and either urbanAEU or urbanBLS is depicted in Figure 2. Finally, the user must provide to urbanSOURCE the concentration measurements d_J $(J = 1, 2, \ldots, N)$, as well as the actual standard deviations σ_J $(J = 1, 2, \ldots, N)$ in the noise process (for Model 1) that define the expected discrepancy between the measured and predicted concentrations or, if these are not known, estimates for the standard deviations s_J $(J = 1, 2, \ldots, N)$ in the noise process (for Model 2). Note that Model 3 does not require the user to input information on the expected uncertainties (either actual or estimated) arising from the noise process.

6 Summary statistics for source parameters

The posterior distribution for Θ provides the full solution for the source reconstruction problem. Inferences on the values of the source parameters are based on this posterior distribution. The posterior distribution may be summarized by various statistics of interest such as the posterior mean of each source parameter, say θ_i (which can be an emission rate Q, or source location $\mathbf{x}_s$, or time at which the source was turned on (off) T_b (T_e)):

$$\overline{\theta_i} = \mathcal{E}[\theta_i | \mathbf{D}] \equiv \int \theta_i p(\Theta | \mathbf{D}, I) \, d\Theta, \tag{34}$$

where $\mathcal{E}[\cdot]$ denotes mathematical expectation. Alternatively, rather than using the posterior mean $\overline{\Theta}$ as the "best" estimate for the source parameter vector Θ, one can use also the *maximum a posteriori* (MAP) estimate of Θ which is defined simply as the mode of the posterior distribution:

$$\widehat{\Theta}_{\mathrm{MAP}} = \underset{\Theta}{\mathrm{argmax}} \; p(\Theta | \mathbf{D}, I). \tag{35}$$

A measure of the uncertainty in the estimate of θ_i can be obtained using the posterior standard deviation $\sigma(\theta_i)$:

$$\begin{aligned} \sigma^2(\theta_i) &= \mathcal{E}[\, (\theta_i - \overline{\theta_i})^2 | \mathbf{D} \,] \\ &\equiv \int (\theta_i - \overline{\theta_i})^2 p(\Theta | \mathbf{D}, I) \, d\Theta. \end{aligned} \tag{36}$$

Alternatively, a $p\%$ credible [or, highest posterior density (HPD)] interval that contains the source parameter θ_i with $p\%$ probability, with the lower and upper bounds of the interval specified such that the probability density within the interval is everywhere larger than that outside it, can be used as a measure of the uncertainty in the determination of θ_i.

In general, Bayesian probability theory never advocates the minimization, maximization, or optimization of any objective or cost function. This is in sharp contrast to the underlying basis for the application of regularization procedures for source reconstruction (see, Thomson et al. [31] and Allen et al. [32], among others) which attempt to find an "optimal" solution by setting a balance between the importance of quality (regularization) and of fitting the data, with sometimes arbitrary choices for the functional used to represent

the regularization in the problem. Rather, the rules of Bayesian probability theory demand that we sum or integrate over unknown quantities, so that the effect is to average over all plausible values of these quantities. The underlying philosophy of Bayesian probability theory for source reconstruction is to find and explore all regions in the hypothesis space (of source distribution models) of reasonably large plausibility, and not simply to find the highest point of maximum posterior probability. This procedure allows a rigorous assessment of the uncertainty in our inferences of the source parameters.

In addition to the summary statistics for the source parameters (e.g., posterior mean, posterior standard deviation, $p\%$ HPD or credible intervals), urbanSOURCE computes two other quantities of interest; namely, (1) the Bayesian evidence Z and (2) the information gain D_{KL}. The Bayesian evidence Z is simply the normalization constant $Z \equiv p(\mathbf{D}|I)$ for the posterior distribution $p(\Theta|\mathbf{D}, I)$; viz.,

$$Z \equiv p(\mathbf{D}|I) = \int p(\Theta|I) p(\mathbf{D}|\Theta, I) \, d\Theta. \tag{37}$$

It it noted that the Bayesian evidence Z (with units of inverse data) can be interpreted as the likelihood of obtaining the given concentration data $\mathbf{D}$, given the background (contextual) information I. The Bayesian evidence is important in the model selection problem, allowing different model assumptions to be compared through ratios of evidence values (or, Bayes factors) [28].

The information gain D_{KL} is a quantitative measure of the gain in information content (about the unknown source distribution) obtained from the receipt of the concentration data $\mathbf{D}$ from the sensor array, resulting from the updating of our state of knowledge concerning the unknown source S (encoded as Θ) from that encoded in the prior distribution $p(\Theta|I)$ to that encoded in the posterior distribution $p(\Theta|\mathbf{D}, I)$. This information gain (amount of useful information about Θ embodied in $\mathbf{D}$) is given by the Kullback-Leibler divergence D_{KL} defined as follows (Cover and Thomas [33]):

$$D_{\mathrm{KL}} \equiv \int \log \left(\frac{p(\Theta|\mathbf{D}, I)}{p(\Theta|I)} \right) p(\Theta|\mathbf{D}, I) \, d\Theta. \tag{38}$$

The Kullback-Leibler divergence defined in Eq. (38) is simply the negative of the entropy (negentropy) of the posterior relative to the prior and, as such, is the information gain provided by the receipt of the concentration data $\mathbf{D}$. More specifically, the information gain "compresses" the posterior relative to the prior so that D_{KL} can simply be interpreted as the logarithm of the volumetric factor by which the prior has been compressed to become the posterior (the greater this compression, the greater is the information gain provided by the concentration data).

In urbanSOURCE, the integrals required in the Bayesian calculations (either in the computation of the various marginal posterior PDFs for the source parameters or in the calculations of the various summary statistics for the parameters and of Z and D_{KL}) are evaluated numerically. The program computes the natural logarithm of the posterior PDF for the source parameter, the reason being that most computers cannot express the large dynamic

range expected in the values of the posterior PDF as one samples the allowable domain of definition for Θ.

7 Application

In this section, we illustrate the application of urbanSOURCE for source reconstruction using two examples. The two examples use real dispersion data sets obtained from the Joint Urban 2003 (JU2003) experiment involving the dispersion of a tracer on an urban-industrial complex scale and from the European Tracer Experiment (ETEX) involving the dispersion of a tracer on a continental scale.

7.1 Joint Urban 2003

The Joint Urban 2003 (JU2003) experiment was an extensive cooperative urban study that was conducted in Oklahoma City, Oklahoma during the period from 28 June 28 to 31 July 31 2003 [34]. The principal objective of JU2003 was to obtain high-quality meteorological and tracer data sets for urban flow and dispersion in a real city environment on a range of scales: namely, from the building (or street) scale encompassing a few buildings or a single street canyon to the neighborhood scale encompassing many city blocks in the central business district (CBD) of Oklahoma City, and finally to the urban scale encompassing the entire CBD and a suburban area several kilometers from downtown Oklahoma City. Furthermore, JU2003 included also measurements of indoor flow and dispersion in four buildings in the CBD, which were coordinated and conducted in conjunction with the outdoor field experiments in order to obtain deeper insights into the physical mechanisms involved in indoor-outdoor exchange rates.

For JU2003, a large number of meteorological measurements in Oklahoma City were undertaken, including detailed measurements of mean wind and turbulence characteristics in the urban canopy and boundary layer obtained with both remote sensing instruments (Doppler sodars and lidars, radar profilers) and fast-response in-situ meteorological sensors (sonic anemometers, infrared thermometers). Additionally, tracer bag samplers were used to measure mean concentration data obtained from the release of a sulfur hexafluoride (SF_6) tracer in downtown Oklahoma City at three different locations. Ten intensive observation periods (IOPs) were undertaken in JU2003, during each of which there were three 30-min (continuous) tracer gas releases. The tracer gas from these releases was sampled in and around downtown Oklahoma City on a regular CBD grid and as far downwind as four kilometers from the release along various sampling arcs.

To illustrate the application of urbanSOURCE, we used concentration data d_J obtained from the second continuous 30-min release of SF_6 in IOP-9, which occurred during the period 06:00–06:30 UTC (01:00-01:30 CDT) on 28 July 2003. The dissemination point for this experiment was located on the south side of Park Avenue (latitude 35.4687° N, longitude 97.5156° W) with a near-surface release height of 1.9 m. The constant gas release

rate for this experiment was $2.0 \, \text{g s}^{-1}$. The highly disturbed mean wind and turbulence fields needed for the calculation of C^* were obtained using urbanSTREAM, which implements a Reynolds-averaged Navier-Stokes approach with a two equation k-ϵ turbulence closure (where k is turbulence kinetic energy, and ϵ is viscous dissipation) [3].

The modeling domain used for the computation of the disturbed wind statistics and for calculation of C^* is shown in Figure 4. The extent of the modeling domain was $1,934.25 \, \text{m} \times 3,610.6 \, \text{m} \times 800.0 \, \text{m}$ in the x- (or, W-E), y- (or, S-N) and z- (or, vertical) directions, respectively, which covers the CBD of Oklahoma City and the surrounding environs. The southwest corner (or, origin) of the modeling domain (see Figure 4) is at the following coordinates in the Universal Transverse Mercator (UTM) coordinate system[4]: zone = 14, $x_0 = 633,683$ UTM easting and $y_0 = 3,923,940$ UTM northing (or, equivalently, in the geodetic coordinate system this location is $35.449959°$ N and $-97.52694°$ E).

The internal coordinate system used in urbanSTREAM is shown in Figure 4, where the southwest corner of the modeling region is chosen as the origin $(0,0)$ in the x-y (horizontal) plane. All distances shown here have been normalized by a reference length scale which is chosen in this case to be $\Lambda_{\text{ref}} = 644.75$ m. Hence, in this internal coordinate system, the northeast corner of the modeling region is referenced as $(3, 5.6)$. A proper subset within this modeling region is chosen as the region in which buildings will be explicitly resolved in the flow simulation; for this example, this rectangular building-aware region $(644.75 \, \text{m} \times 709.23$ m) has its southwest corner at $(1, 2.5)$ and its northeast corner at $(2, 3.6)$. In the portion of the modeling region lying outside the building-aware region, all buildings are treated as virtual and their effects on the flow are modeled using a distributed drag force representation in the mean momentum equations. Figure 4 also shows that the location of the tracer source (blue dot) was at $(x_s, y_s) = (1.5537, 3.2506)$ (in the normalized local coordinate system used by the model) or at $(x_s, y_s) = (1001.7, 2095.7)$ m (in the un-normalized coordinate system). In other words, the source was located 1001.7 m east and 2095.7 m north of the origin $(0,0)$ shown in Figure 4.

A mesh of $99 \times 139 \times 69$ grid lines in the x-, y-, and z-directions, respectively, was used to accommodate all the necessary geometrical details. The interior building-aware region was covered with a fine calculation grid of $55 \times 100 \times 69$ grid lines to better approximate the building features in this region. The fine grid used for the building-aware region contains 379,500 nodes, whereas the entire computational domain was covered with a mesh of 945,509 nodes. The grid lines were preferentially concentrated near the solid surfaces (ground, building rooftops and walls) where the gradients in the flow properties are expected to be greatest, and the spacing between the grid lines was gently stretched with increasing distance from the solid surfaces.

The velocity statistics, computed over the computational domain shown in Figure 4, were used in urbanAEU to calculate C^* for a number of tracer bag samplers located in the domain. In urbanAEU, the turbulent diffusivity K is obtained from the turbulent (eddy)

---

[4]The UTM easting coordinate reported here is referenced relative to the central meridian of the zone.
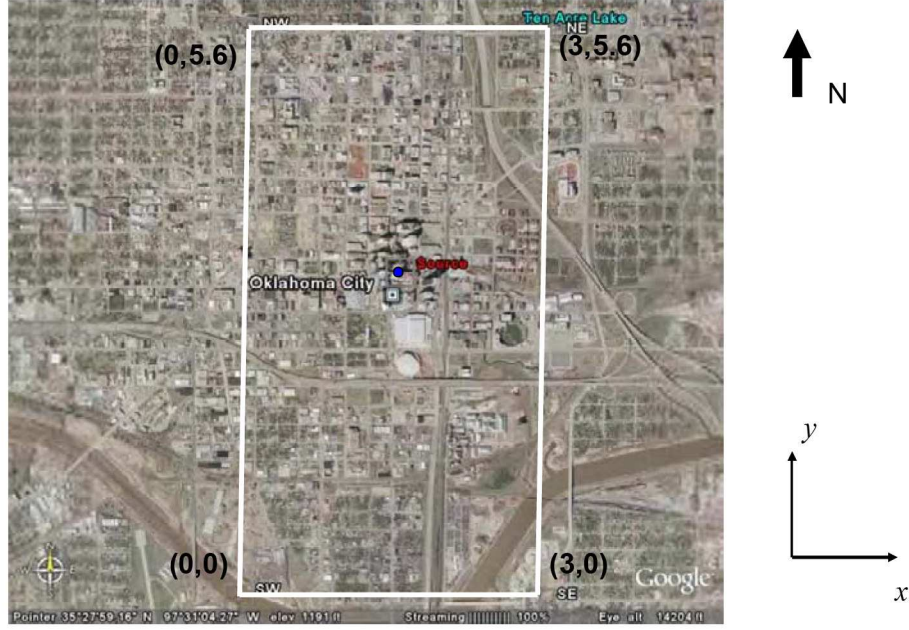
Figure 4: *The computational domain used for the prediction of the disturbed wind statistics in the CBD of Oklahoma City, as well as for computation of C^*. The location of the source is indicated using a blue dot and lies near the centre of the computational domain.*

viscosity (predicted by urbanSTREAM) in combination with a turbulent Schmidt number $Sc_t \approx 0.63$ in the following manner: $K = \nu_t / Sc_t$ [cf. Eq. (6)]. Now, we use C^* and concentration data d_J obtained from the bag samplers in urbanSOURCE to reconstruct the (assumed here to be unknown) source characteristics. Because we are dealing with a continuous source, the only relevant parameters here are the source location $\mathbf{x}_s$ and the emission rate Q. Furthermore, since we are dealing with near-surface sources with $z_s \approx 0$, the only relevant unknown location parameters for the unknown source are its W-E (x_s) and S-N (y_s) positions (relative to the fixed origin $(0,0)$ of the computational domain displayed in Figure 4).

We consider three different cases involving the use of different numbers of detectors in the array for source inversion. In the tracer field experiment from which the concentration data were extracted, the prevailing winds for the experiment were from the south at about 6.8 m s^{-1} at 50 metres above ground level at the southern edge of the computational domain shown in Figure 4. For each of the cases used to test the source inversion, the parameters that define the prior distribution $p(\Theta|I)$ [cf. Eq. (23)] are chosen as follows: $Q_{\max} = 10.0$ g s^{-1}; and, $\mathcal{D} = [0.0, 1934.25]$ m $\times$ [0.0, 3610.6] m (or, equivalently, $\mathcal{D} = [0.0, 3.0] \times [0.0, 5.6]$ in the normalized local coordinate system used by urbanSTREAM) providing the prior bounds on the source location in the (x, y)-plane. Recall that the (x, y) coordinate system used here is chosen as shown in Figure 4. Finally, for Model 2, the hyperparameter α was set to the default value (in urbanSOURCE) of $\alpha = \pi^{-1}$.
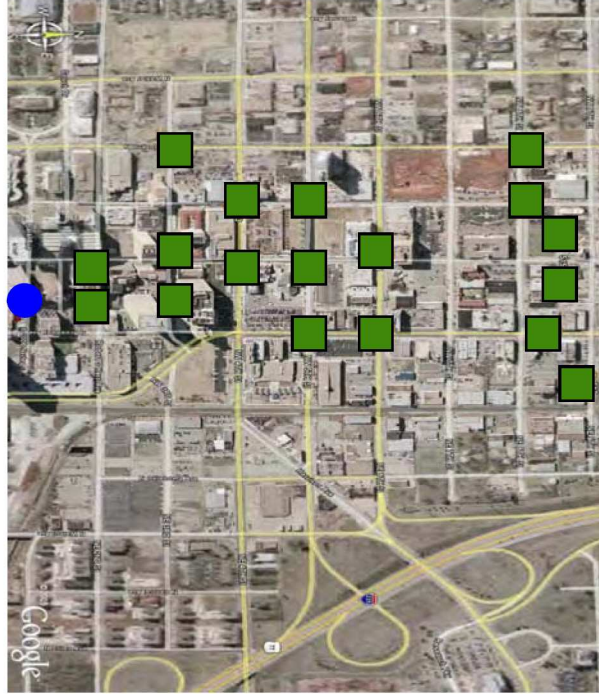
Figure 5: *Case 1: source reconstruction using 18 detectors. The solid blue dot shows the location of the source. The filled solid green squares mark the location of the detectors in the CBD (of Oklahoma City) that were used for source reconstruction.*

7.1.1 Case 1: 18 detectors

In case 1, the detectors used for the source reconstruction algorithm are shown in Figure 5. As can be seen from this figure, this case involves the use of 18 detectors in the array for source inversion. The reconstruction of the source parameters for this case was undertaken using all three models for the posterior distribution [cf. Eqs. (31), (32) and (33)]. In Model 1, the standard deviations σ_J of the expected discrepancies between the measured concentrations d_J and the predicted concentrations $\overline{C}_J$ (accounting for input, model and measurement errors) were obtained from a previous model validation study [3] undertaken to evaluate the predictive accuracy of urbanEU/urbanAEU. In contrast, for Model 2, only very crude (and certainly incorrect) estimates s_J for the uncertainties σ_J were used; namely, the estimates of the uncertainties for the two largest values of the measured concentration were simply set to be equal to 10% of the measured concentration and the estimates of the uncertainties for all other values of the measured concentration were set to be equal to 25% of the measured concentration. Recall that for Model 2, our poor knowledge of the
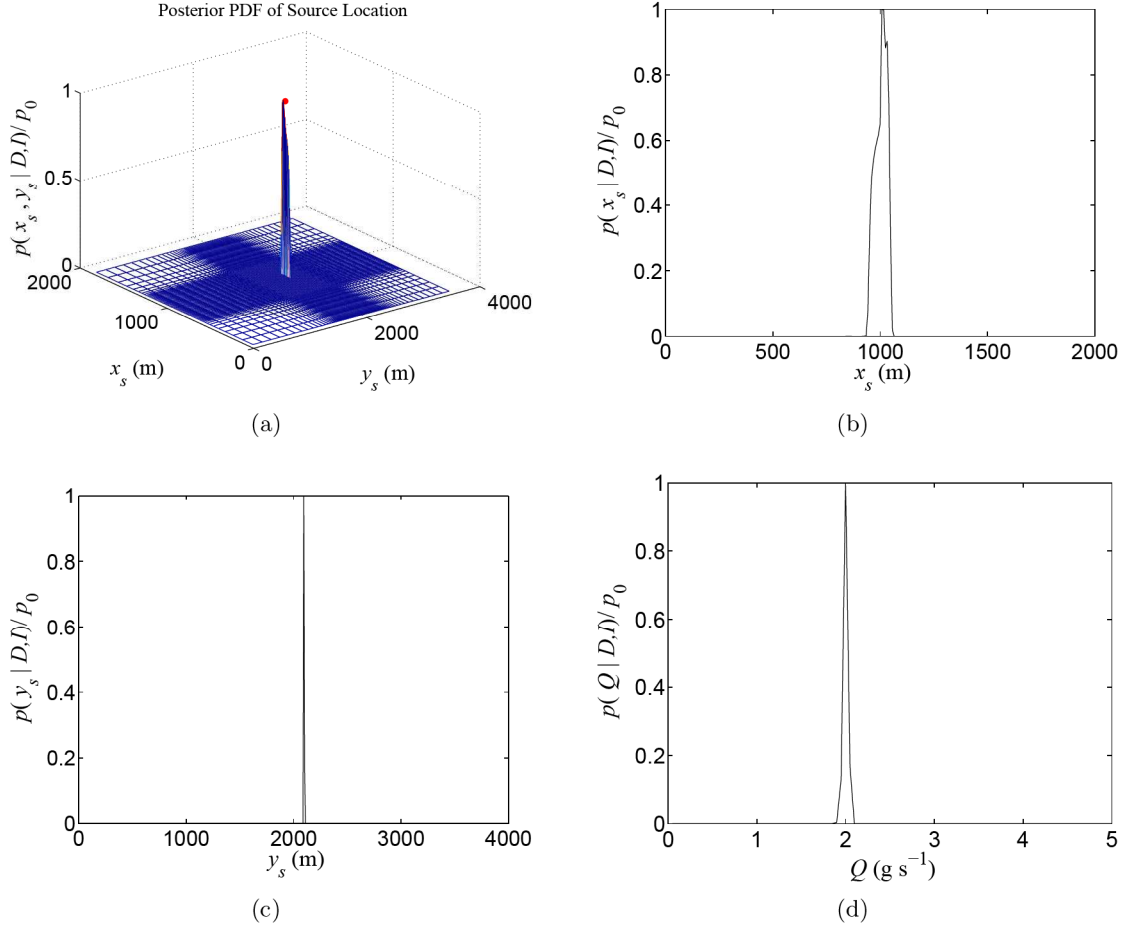
Figure 6: *The marginal joint posterior PDF of source location $p(x_s, y_s | \mathbf{D}, I)$ [upper left panel], the marginal posterior PDF of W-E source location $p(x_s | \mathbf{D}, I)$ [upper right panel], the marginal posterior PDF of S-N source location $p(y_s | \mathbf{D}, I)$ [lower left panel], and the marginal posterior PDF of the emission rate $p(Q | \mathbf{D}, I)$ [lower right panel] obtained from Model 1 for source reconstruction using 18 concentration detectors. All PDFs have been normalized by their maximum value p_0. The true source location is indicated using the red dot in the upper left panel.*

uncertainties σ_J (embodied in the estimates s_J) is explicitly accounted for by incorporating a distribution for the true (but unknown) uncertainties.

The posterior distribution $p(\Theta | \mathbf{D}, I)$ was directly evaluated using urbanSOURCE. The resulting marginal joint posterior PDF for the source location (x_s, y_s), the marginal posterior PDF for the W-E coordinate x_s of the source location, the marginal posterior PDF for the S-N coordinate y_s of the source location, and the marginal posterior PDF of the emission rate Q are shown in Figure 6 for Model 1. Furthermore, the posterior mean, MAP estimate, posterior standard deviation, and lower and upper bounds for the 97.5% HPD interval of

the source parameters (x_s, y_s, Q) are summarized in Table 1.

Comparing the estimated values of the source parameters with the actual values of the source parameters, it is seen that the algorithm using Model 1 has enabled the recovery of the true parameters to within the stated errors. In particular, the actual location of the source was correctly estimated in this case to within a one standard deviation interval, with the S-N (y_s) position of the source determined with an accuracy (± 2.7 m) that was about ten times greater than the accuracy (± 27.5 m) for the determination of the W-E (x_s) position of the source. Finally, Table 1 indicates that the information gain obtained from the concentration $\mathbf{D}$ (and from our knowledge of the uncertainty in this data) was found to be $D_{\mathrm{KL}} = 13.6$ natural units (nits), implying that the information contained in the concentration data allowed the "posterior volume" of the hypothesis space (volume of hypothesis space of reasonably large plausibility after receipt of the concentration data) to decrease by a factor of $\exp(D_{\mathrm{KL}}) \approx 8.1 \times 10^5$ relative to the "prior volume" of the hypothesis space (volume of hypothesis space of reasonably large plausibility before the receipt of the concentration data).

The marginal posterior PDFs for the source location and emission rate using Model 2 with the 18 concentration detectors (shown in Figure 5) are presented in Figure 7. In addition, various summary statistics for the source parameters obtained from the marginal posterior PDFs in Figure 7 are summarized in Table 2. A comparison of Figure 6 with Figure 7 shows that the "characteristic widths" of the marginal posterior PDFs for the source location and emission rate obtained from Model 2 are broader than those obtained from Model 1 (implying a greater uncertainty in the recovery of the source parameters). Nevertheless, the estimates for the source parameters obtained using Model 2 are still very good – the best estimates (based on the posterior mean) for the x_s and y_s locations of the source were 1003.0 ± 42.9 m and 2108.7 ± 25.1 m, respectively, and for the emission rate Q was 1.83 ± 0.18 g s^{-1} (with accuracy estimates at one standard deviation). Again, the location and emission rate of the source was correctly estimated using Model 2 to within a one standard deviation interval. However, the uncertainties in the determination of x_s, y_s and Q using Model 2

Table 1: *The posterior mean, maximum a posteriori (MAP) estimate, posterior standard deviation, and lower and upper bounds of the 97.5% HPD interval of the parameters x_s (m), y_s (m), and Q (g s^{-1}) obtained for Model 1 using 18 concentration detectors for source inversion. The information gain D_{KL} (measured in natural units or nits) obtained from the concentration data is summarized in the last row of the table.*

| Parameter	Mean	MAP	Standard Deviation	97.5% HPD	Actual
x_s (m)	1003.8	1024.2	27.5	$(950.8, 1048.7)$	1001.7
y_s (m)	2095.8	2094.7	2.7	$(2094.7, 2102.5)$	2095.7
Q (g s^{-1})	2.00	2.00	0.03	$(1.95, 2.05)$	2.00
$D_{\mathrm{KL}} = 13.6$ nits					

Figure 7: *The marginal joint posterior PDF of source location $p(x_s, y_s | \mathbf{D}, I)$ [upper left panel], the marginal posterior PDF of W-E source location $p(x_s | \mathbf{D}, I)$ [upper right panel], the marginal posterior PDF of S-N source location $p(y_s | \mathbf{D}, I)$ [lower left panel], and the marginal posterior PDF of the emission rate $p(Q | \mathbf{D}, I)$ [lower right panel] obtained from Model 2 for source reconstruction using 18 concentration detectors. All PDFs have been normalized by their maximum value p_0. The true source location is indicated using the red dot in the upper left panel.*

were about 1.5, 10, and 6 times larger, respectively, than the corresponding uncertainties obtained using Model 1. Finally, the information gain for Model 2 was $D_{\mathrm{KL}} = 10.7$ nits, which is about 3 nits less than that for Model 1.

The posterior PDFs for the source location and the emission rate for Model 3 with 18 concentration detectors are displayed in Figure 8. As is evident by comparison of this figure with Figures 6 and 7, the probability distributions $p(x_s, y_s | \mathbf{D}, I)$, $p(x_s | \mathbf{D}, I)$ and $p(y_s | \mathbf{D}, I)$ do not have a typical Gaussian-like form with the maxima located (approximately or better) at the actual source location; rather, the posterior distributions for the source location for Model 3 is multimodal, exhibiting a cluster of peaks (modes) that are centred

Table 2: *The posterior mean, maximum a posteriori (MAP) estimate, posterior standard deviation, and lower and upper bounds of the 97.5% HPD interval of the parameters x_s (m), y_s (m), and Q (g s^{-1}) obtained for Model 2 using 18 concentration detectors for source inversion. The information gain D_{KL} (measured in natural units or nits) obtained from the concentration data is summarized in the last row of the table.*

Parameter	Mean	MAP	Standard Deviation	97.5% HPD	Actual
x_s (m)	1003.0	1032.4	42.9	$(934.4, 1089.5)$	1001.7
y_s (m)	2108.7	2149.4	25.1	$(2094.7, 2180.7)$	2095.7
Q (g s^{-1})	1.83	1.85	0.18	$(1.15, 2.15)$	2.00
$D_{\mathrm{KL}} = 10.7$ nits					

(approximately or better) at the actual source location. Furthermore, it is seen that the posterior PDF for emission rate for Model 3 is bimodal in form — there is a maximum (mode) at 2.00 g s^{-1} that coincides with the true emission rate and a secondary maximum (mode) at about 3.5 g s^{-1}.

Table 3 summarizes the posterior mean, MAP estimate, posterior standard deviation, and lower and upper bounds for the 97.5% HPD interval of the source parameters obtained by numerical integration of the posterior PDF calculated from Model 3. From this information, we see that the parameters of the unknown source have been characterized to a reasonable accuracy, although it is seen that the precision in the determination of the source parameters using Model 3 is worse than that of Models 1 and 2. This should not be too surprising owing to the fact that Model 3 for the posterior distribution of Θ does not contain information on either the actual (as in Model 1) or estimated (as in Model 2) uncertainties (input, model and measurement) associated with each concentration datum d_J $(J = 1, 2, \ldots, N)$.

For Model 3, the location of the source was estimated to be $\widehat{x}_s = 992.8 \pm 51.7$ m, $\widehat{y}_s = 1985.5 \pm 159.0$ m and $\widehat{Q} = 2.69 \pm 0.77$ g s^{-1} (with precision estimates at one standard deviation). Although the precision in the determination of x_s is comparable to that in Model 2, the precision of the estimates of y_s and Q are about 6 and 4 times worse, respectively, than those obtained in Model 2. Obviously, even crude estimates for the uncertainties associated with the concentration (measured and predicted) is advantageous for the source reconstruction. However, in spite of the poorer estimates in the source parameters when compared to Model 2, the larger uncertainty bounds obtained for Model 3 nevertheless ensure that all the source parameters are correctly estimated to within one standard deviation interval.

Finally, for Model 3, the information gain provided by the concentration data **D** was found to be only $D_{\mathrm{KL}} = 8.6$ nits, which is about 5 and 2 nits less information gain than obtained using Models 1 and 2 for source reconstruction, respectively. The difference in information gain here resides in the concentration data uncertainties used in Models 1 and 2.
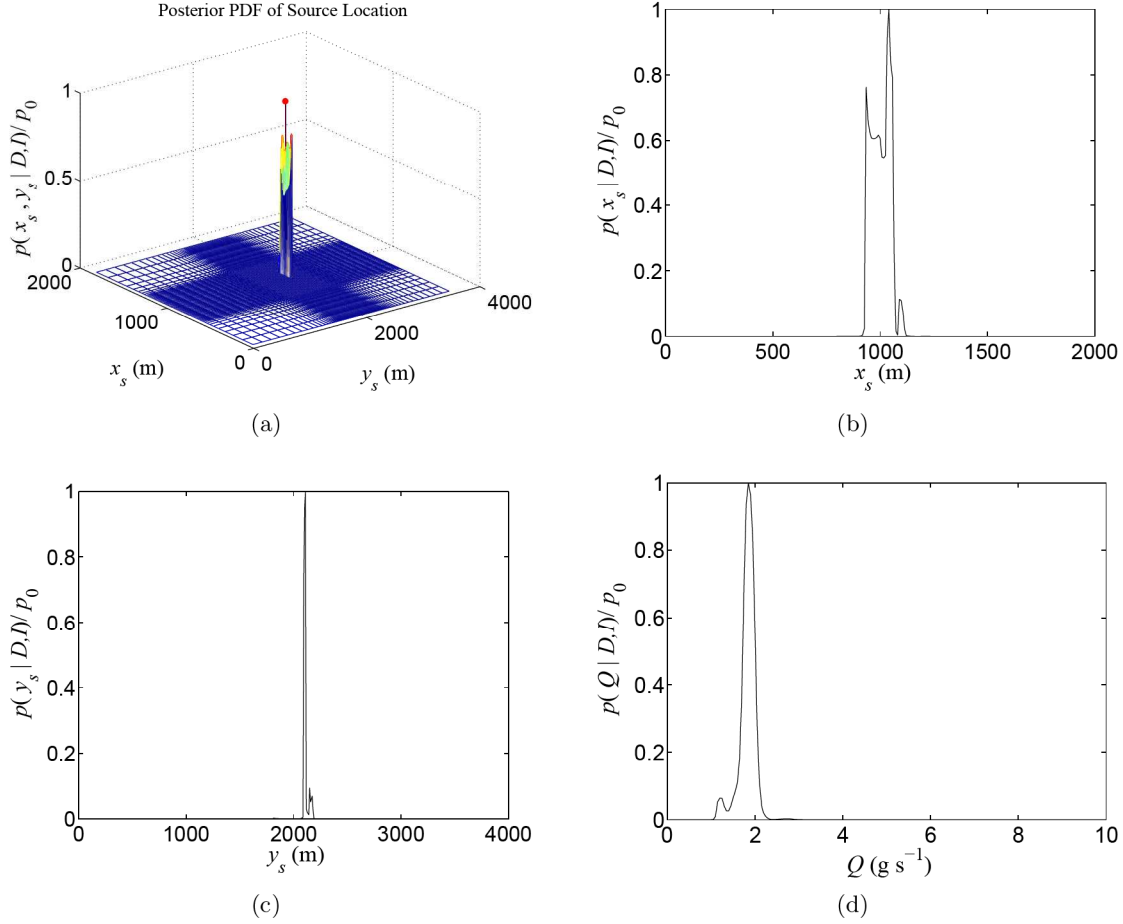
Figure 8: *The marginal joint posterior PDF of source location $p(x_s, y_s | \mathbf{D}, I)$ [upper left panel], the marginal posterior PDF of W-E source location $p(x_s | \mathbf{D}, I)$ [upper right panel], the marginal posterior PDF of S-N source location $p(y_s | \mathbf{D}, I)$ [lower left panel], and the marginal posterior PDF of the emission rate $p(Q | \mathbf{D}, I)$ [lower right panel] obtained from Model 3 for source reconstruction using 18 concentration detectors. All PDFs have been normalized by their maximum value p_0. The true source location is indicated using the red dot in the upper left panel.*

Table 3: *The posterior mean, maximum a posteriori (MAP) estimate, posterior standard deviation, and lower and upper bounds of the 97.5% HPD interval of the parameters x_s (m), y_s (m), and Q (g s^{-1}) obtained for Model 3 using 18 concentration detectors for source inversion. The information gain D_{KL} (measured in natural units or nits) obtained from the concentration data is summarized in the last row of the table.*

Parameter	Mean	MAP	Standard Deviation	97.5% HPD	Actual
x_s (m)	992.8	926.3	51.7	$(893.6, 1097.7)$	1001.7
y_s (m)	1985.5	2094.7	159.0	$(1813.4, 2188.5)$	2095.7
Q (g s^{-1})	2.69	2.00	0.77	$(1.40, 4.3)$	2.00
$D_{\mathrm{KL}} = 8.6$ nits					

Figure 9: Case 2: source reconstruction using 9 detectors. The solid blue dot shows the location of the source. The filled solid green squares mark the location of the detectors in the CBD (of Oklahoma City) that were used for source reconstruction.

7.1.2 Case 2: 9 detectors

To further test the source reconstruction algorithm, we use only 9 concentration data (a subset that consisted of only 50% of the data used in Case 1). The locations of the bag samplers which provided the 9 concentration data are displayed in Figure 9. As in Case 1, the source reconstruction was undertaken using all three models for the posterior distribution $p(\Theta|\mathbf{D}, I)$ [cf. Eqs. (31), (32) and (33)].

Figures 10, 11 and 12 shows the joint marginal posterior distribution for (x_s, y_s), as well as the marginal posterior distributions for x_s, y_s, and Q obtained for Case 2 using Models 1, 2 and 3, respectively, for $p(\Theta|\mathbf{D}, I)$. In addition, summary statistics (posterior mean, MAP estimate, posterior standard deviation, lower and upper bounds of the 97.5% HPD intervals) for the source parameters as well as the information gain D_{KL} of Models 1, 2, and 3 are tabulated in Tables 4, 5 and 6, respectively.

Even with only 9 concentration detectors, it is seen (cf. Figure 10 and Table 4) that the source location and emission rate have been determined with very good accuracy using Model 1. The mean and standard deviation estimates of x_s, y_s and Q using Model 1 (for which accurate estimates of the uncertainties σ_J are assumed to be available) are as follows: $\widehat{x}_s = 1005.1 \pm 26.9$ m, $\widehat{y}_s = 2097.8 \pm 12.2$ m and $\widehat{Q} = 1.99 \pm 0.04$ g s^{-1}. Note that the estimates are very comparable to those obtained using Model 1 in Case 1 with 18 detectors. The largest difference lies in the uncertainty in the determination of y_s. In particular, using the 9 detectors shown in Figure 9 rather than 18 detectors displayed in Figure 5, the precision in the determination of y_s has decreased by a factor of about 4.5.

Using only crude estimates for the uncertainties in the concentration data results in a deterioration in estimates for the source parameters, as can be seen by comparison of the results of the source reconstruction for Model 2 (cf. Figure 11 and Table 5) with those for Model 1 (cf. Figure 10 and Table 4). Note that although the source parameter estimates
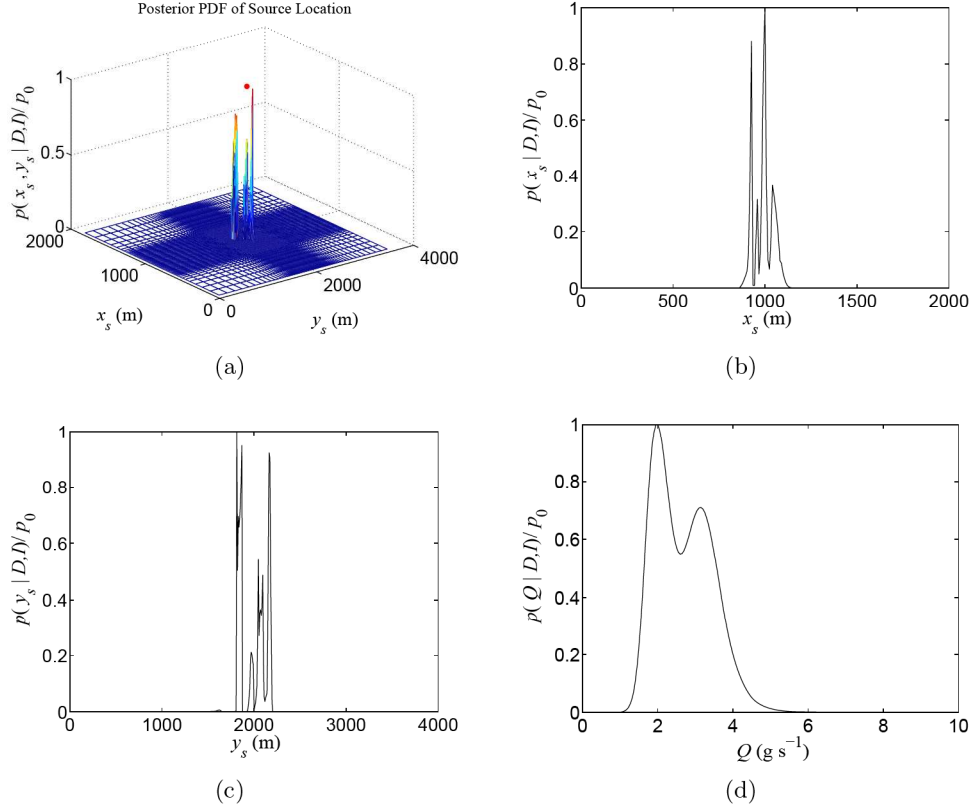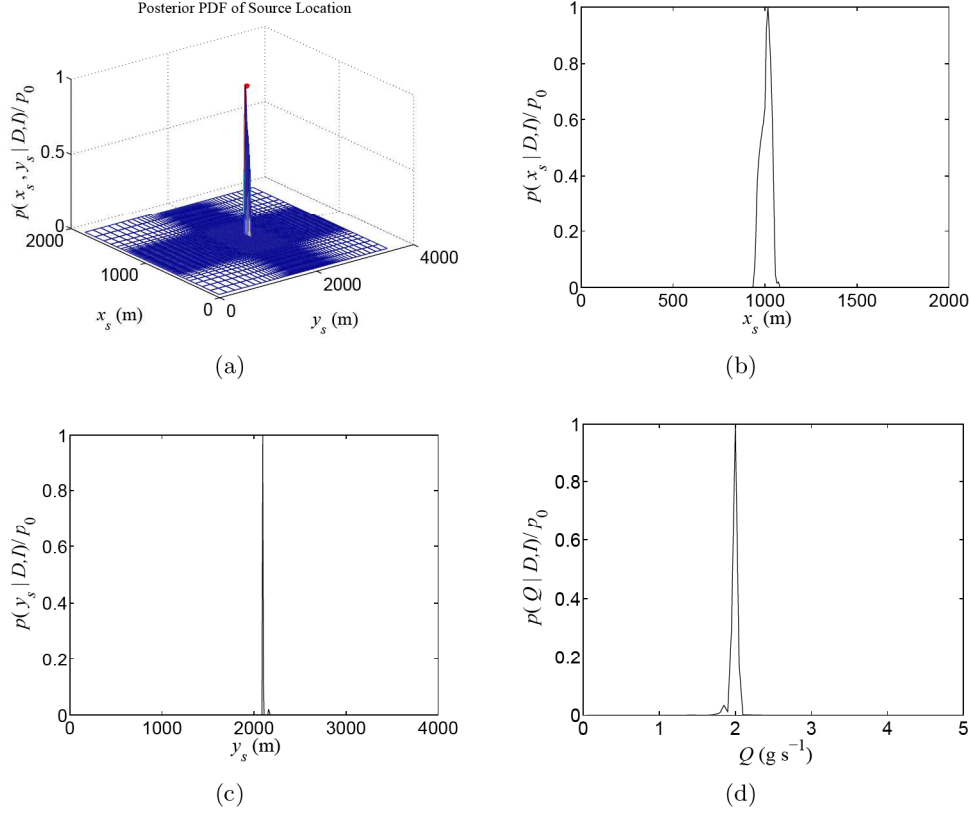
Figure 10: *The marginal joint posterior PDF of source location $p(x_s, y_s|\mathbf{D}, I)$ [upper left panel], the marginal posterior PDF of W-E source location $p(x_s|\mathbf{D}, I)$ [upper right panel], the marginal posterior PDF of S-N source location $p(y_s|\mathbf{D}, I)$ [lower left panel], and the marginal posterior PDF of the emission rate $p(Q|\mathbf{D}, I)$ [lower right panel] obtained from Model 1 for source reconstruction using 9 concentration detectors. All PDFs have been normalized by their maximum value p_0. The true source location is indicated using the red dot in the upper left panel.*

Table 4: *The posterior mean, maximum a posteriori (MAP) estimate, posterior standard deviation, and lower and upper bounds of the 97.5% HPD interval of the parameters x_s (m), y_s (m), and Q (g s^{-1}) obtained for Model 1 using 9 concentration detectors for source inversion. The information gain D_{KL} (measured in natural units or nits) obtained from the concentration data is summarized in the last row of the table.*

Parameter	Mean	MAP	Standard Deviation	97.5% HPD	Actual
x_s (m)	1005.1	1016.0	26.9	$(950.8, 1048.7)$	1001.7
y_s (m)	2097.8	2094.7	12.2	$(2094.7, 2157.2)$	2095.7
Q (g s^{-1})	1.99	2.00	0.04	$(1.85, 2.05)$	2.00
$D_{\mathrm{KL}} = 13.2$ nits					

Posterior PDF of Source Location

(a)

(b)

(c)

(d)

Figure 11: *The marginal joint posterior PDF of source location $p(x_s, y_s|\mathbf{D}, I)$ [upper left panel], the marginal posterior PDF of W-E source location $p(x_s|\mathbf{D}, I)$ [upper right panel], the marginal posterior PDF of S-N source location $p(y_s|\mathbf{D}, I)$ [lower left panel], and the marginal posterior PDF of the emission rate $p(Q|\mathbf{D}, I)$ [lower right panel] obtained from Model 2 for source reconstruction using 9 concentration detectors. All PDFs have been normalized by their maximum value p_0. The true source location is indicated using the red dot in the upper left panel.*

Table 5: *The posterior mean, maximum a posteriori (MAP) estimate, posterior standard deviation, and lower and upper bounds of the 97.5% HPD interval of the parameters x_s (m), y_s (m), and Q (g s^{-1}) obtained for Model 2 using 9 concentration detectors for source inversion. The information gain D_{KL} (measured in natural units or nits) obtained from the concentration data is summarized in the last row of the table.*

| Parameter	Mean	MAP	Standard Deviation	97.5% HPD	Actual
x_s (m)	1058.9	1065.0	47.5	$(942.6, 1122.1)$	1001.7
y_s (m)	2105.9	2102.5	103.5	$(2071.3, 2172.9)$	2095.7
Q (g s^{-1})	1.68	1.3	0.38	$(1.25, 2.10)$	2.00
$D_{\mathrm{KL}} = 12.1$ nits					

Figure 12: *The marginal joint posterior PDF of source location $p(x_s, y_s | \mathbf{D}, I)$ [upper left panel], the marginal posterior PDF of W-E source location $p(x_s | \mathbf{D}, I)$ [upper right panel], the marginal posterior PDF of S-N source location $p(y_s | \mathbf{D}, I)$ [lower left panel], and the marginal posterior PDF of the emission rate $p(Q | \mathbf{D}, I)$ [lower right panel] obtained from Model 3 for source reconstruction using 9 concentration detectors. All PDFs have been normalized by their maximum value p_0. The true source location is indicated using the red dot in the upper left panel.*

Table 6: *The posterior mean, maximum a posteriori (MAP) estimate, posterior standard deviation, and lower and upper bounds of the 97.5% HPD interval of the parameters x_s (m), y_s (m), and Q (g s^{-1}) obtained for Model 3 using 9 concentration detectors for source inversion. The information gain D_{KL} (measured in natural units or nits) obtained from the concentration data is summarized in the last row of the table.*

Parameter	Mean	MAP	Standard Deviation	97.5% HPD	Actual
x_s (m)	1055.7	1065.0	54.6	(934.4, 1138.5)	1001.7
y_s (m)	1617.8	2102.5	632.4	(246.1, 2204.1)	2095.7
Q (g s^{-1})	3.68	1.65	2.36	(1.15, 9.10)	2.00
$D_{\mathrm{KL}} = 5.7$ nits					
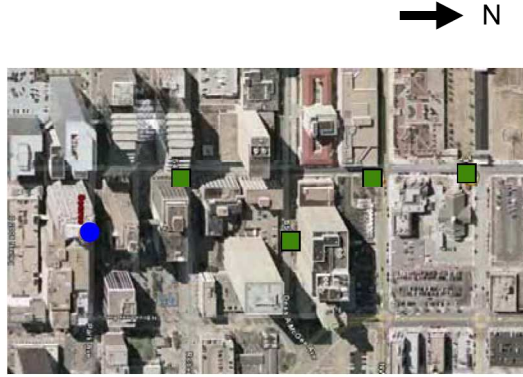
Figure 13: *Case 3: source reconstruction using 4 detectors. The solid blue dot shows the location of the source. The filled solid green squares mark the location of the detectors in the CBD (of Oklahoma City) that were used for source reconstruction.*

for Model 2 are still quite good, the uncertainties in the determination of x_s, y_s and Q are larger than those for Model 1 by roughly factors of 1.8, 8.5 and 10, respectively.

As in Case 1, the estimated values of the source parameters are generally quite poorly determined using Model 3 with only 9 concentration detectors. In this example, only the W-E location x_s of the source is relatively well determined: $\widehat{x}_s = 1055.7 \pm 54.6$ m. However, the information is not sufficient to estimate the S-N location y_s and emission rate Q with good precision. Indeed, the mean and standard deviation estimates of y_s and Q provided by Model 3 in this example are $\widehat{y}_s = 1617.8 \pm 632.4$ m and $\widehat{Q} = 3.68 \pm 2.36$ g s^{-1}. The absolute percentage discrepancies between the true and estimated values of y_s and Q are 23% and 84%, respectively, but the quoted uncertainty bounds here are nevertheless large enough to ensure that the estimated values for these source parameters have been adequately recovered to within the stated errors. Once more, it can be seen that information concerning expected uncertainties in the measured and predicted concentration (even when crudely specified as in Model 2) are useful for improving the accuracy and precision of the source reconstruction.

7.1.3 Case 3: 4 detectors

In case 3, the detectors used for the source reconstruction algorithm are shown in Figure 13. As can be seen from this figure, this difficult case involves the use of only 4 detectors for the source inversion (a subset that represents less than 25% of the detectors used in case 1). The problem is made more difficult by the fact that only one of the detectors lies at or near the plume centreline. The source reconstruction was undertaken in this case using all three models for the posterior distribution $p(\Theta|\mathbf{D}, I)$. The results for the source reconstruction in the form of marginal posterior PDFs for the source parameters are summarized in Figures 14, 15 and 16 for Models 1, 2 and 3, respectively. The summary statistics for x_s, y_s and Q obtained from the marginal posterior PDFs which include the posterior mean, MAP estimate (or posterior mode), posterior standard deviation, and the
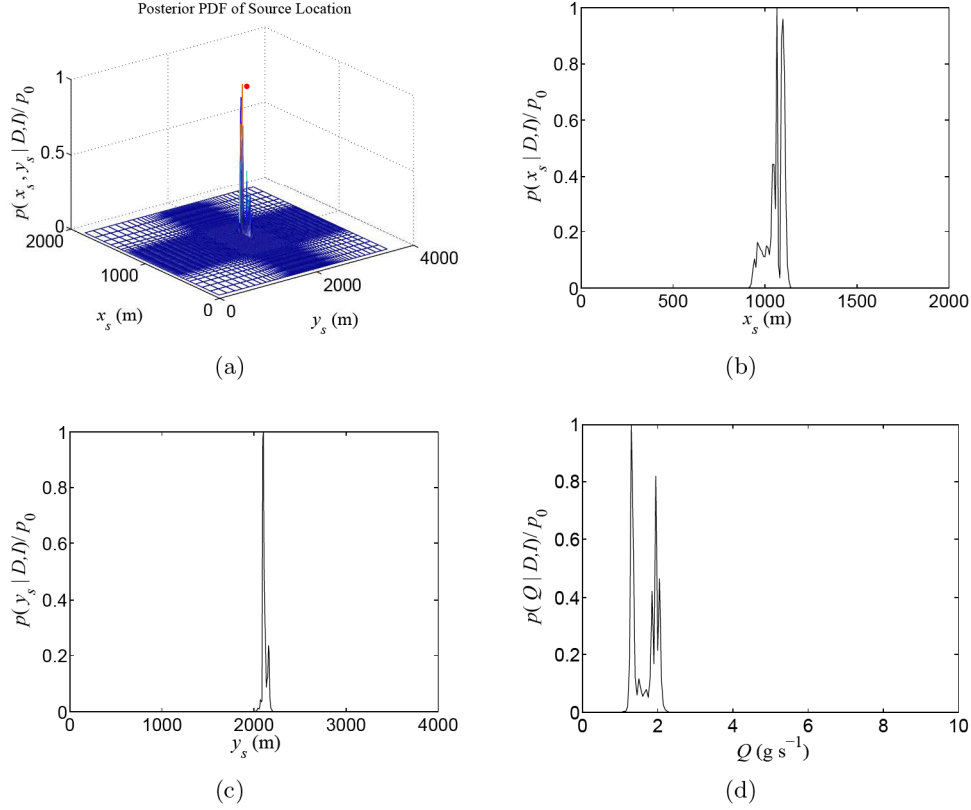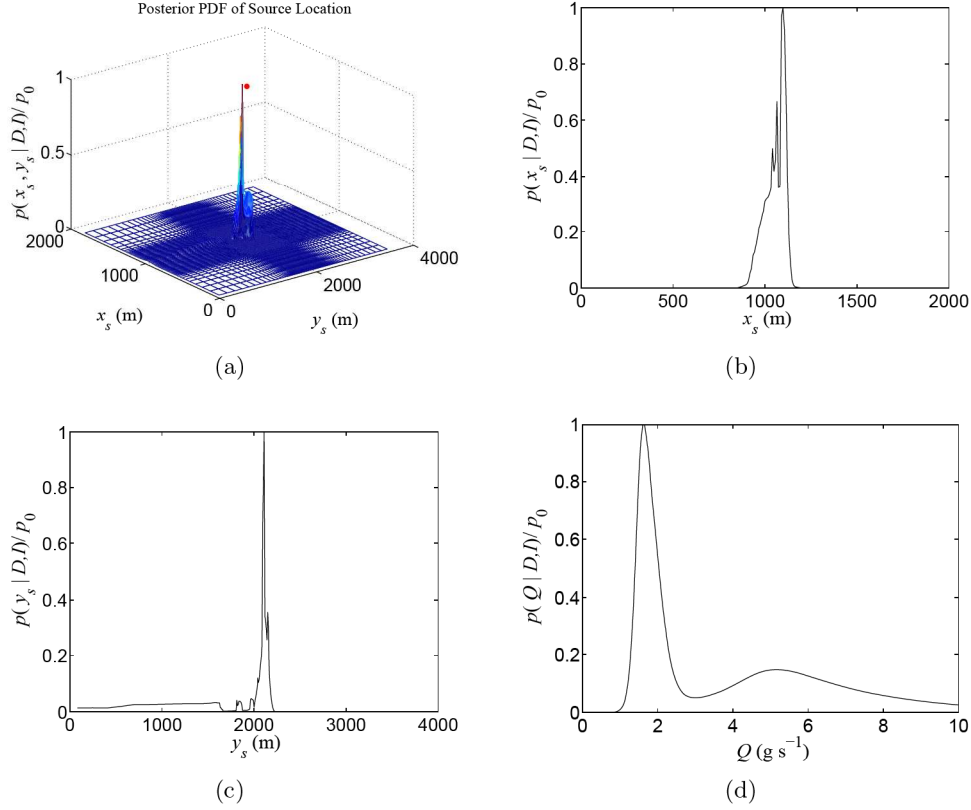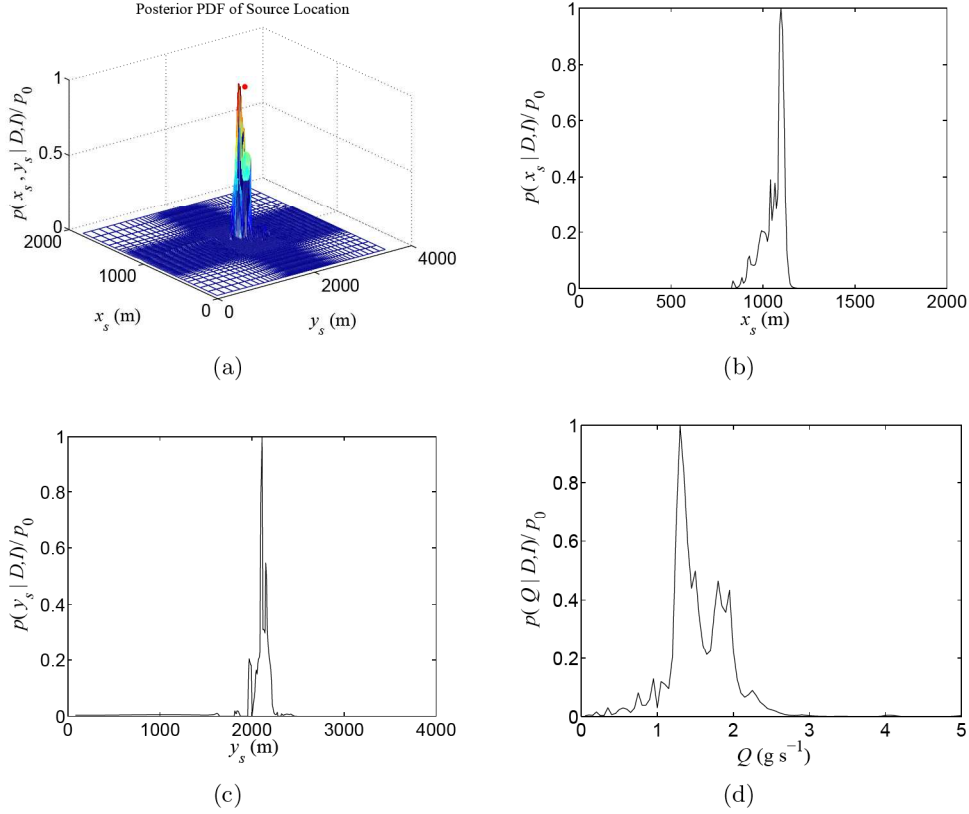
Figure 14: *The marginal joint posterior PDF of source location $p(x_s, y_s|\mathbf{D}, I)$ [upper left panel], the marginal posterior PDF of W-E source location $p(x_s|\mathbf{D}, I)$ [upper right panel], the marginal posterior PDF of S-N source location $p(y_s|\mathbf{D}, I)$ [lower left panel], and the marginal posterior PDF of the emission rate $p(Q|\mathbf{D}, I)$ [lower right panel] obtained from Model 1 for source reconstruction using 4 concentration detectors. All PDFs have been normalized by their maximum value p_0. The true source location is indicated using the red dot in the upper left panel.*

Table 7: *The posterior mean, maximum a posteriori (MAP) estimate, posterior standard deviation, and lower and upper bounds of the 97.5% HPD interval of the parameters x_s (m), y_s (m), and Q (g s^{-1}) obtained for Model 1 using 4 concentration detectors for source inversion. The information gain D_{KL} (measured in natural units or nits) obtained from the concentration data is summarized in the last row of the table.*

Parameter	Mean	MAP	Standard Deviation	97.5% HPD	Actual
x_s (m)	1058.4	1089.5	59.4	$(910.0, 1138.5)$	1001.7
y_s (m)	1988.4	2102.5	395.0	$(1618.0, 2274.5)$	2095.7
Q (g s^{-1})	1.54	1.3	0.41	$(0.35, 2.45)$	2.00
$D_{\mathrm{KL}} = 10.1$ nits					
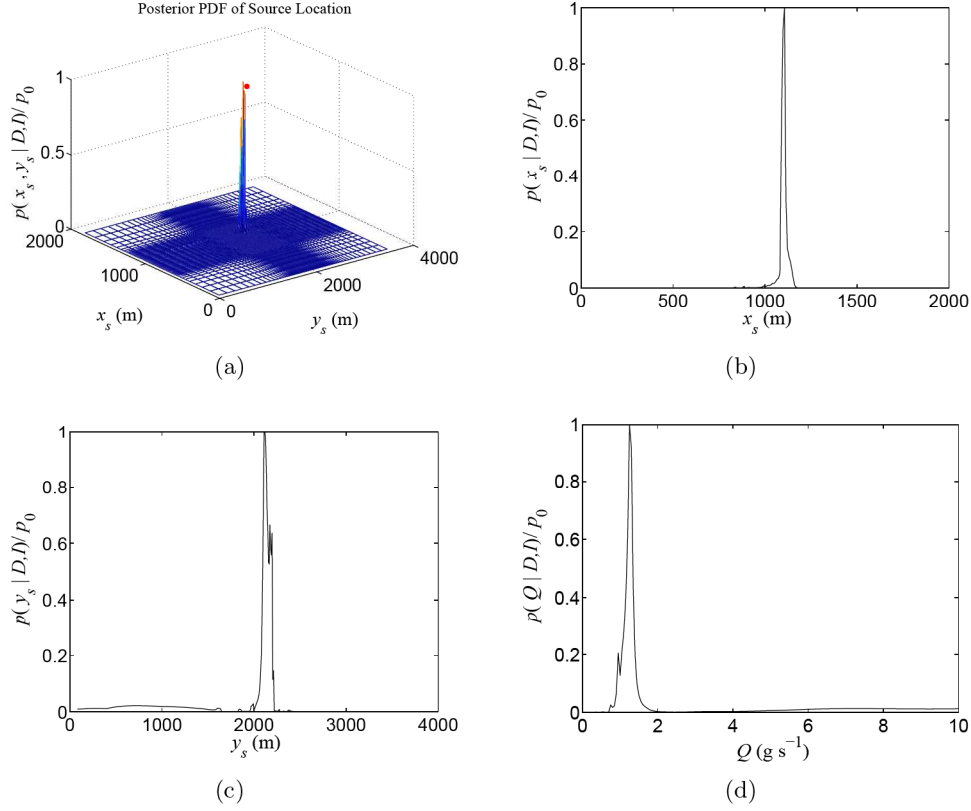
Figure 15: *The marginal joint posterior PDF of source location $p(x_s, y_s|\mathbf{D}, I)$ [upper left panel], the marginal posterior PDF of W-E source location $p(x_s|\mathbf{D}, I)$ [upper right panel], the marginal posterior PDF of S-N source location $p(y_s|\mathbf{D}, I)$ [lower left panel], and the marginal posterior PDF of the emission rate $p(Q|\mathbf{D}, I)$ [lower right panel] obtained from Model 2 for source reconstruction using 4 concentration detectors. All PDFs have been normalized by their maximum value p_0. The true source location is indicated using the red dot in the upper left panel.*

Table 8: *The posterior mean, maximum a posteriori (MAP) estimate, posterior standard deviation, and lower and upper bounds of the 97.5% HPD interval of the parameters x_s (m), y_s (m), and Q (g s^{-1}) obtained for Model 2 using 4 concentration detectors for source inversion. The information gain D_{KL} (measured in natural units or nits) obtained from the concentration data is summarized in the last row of the table.*

| Parameter	Mean	MAP	Standard Deviation	97.5% HPD	Actual
x_s (m)	1100.0	1105.8	28.7	$(1032.3, 1154.8)$	1001.7
y_s (m)	1858.8	2172.9	567.0	$(556.9, 2211.9)$	2095.7
Q (g s^{-1})	2.54	1.25	2.61	$(0.75, 10.0)$	2.00
$D_{KL} = 9.4$ nits					
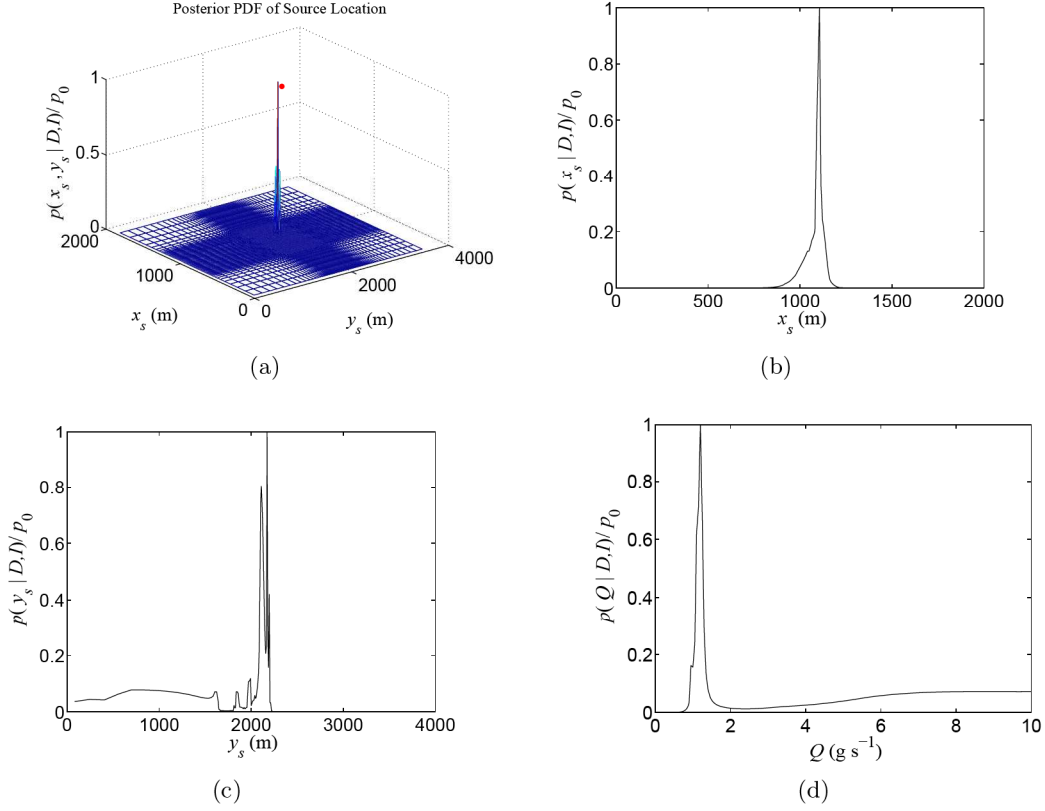
Figure 16: *The marginal joint posterior PDF of source location $p(x_s, y_s | \mathbf{D}, I)$ [upper left panel], the marginal posterior PDF of W-E source location $p(x_s | \mathbf{D}, I)$ [upper right panel], the marginal posterior PDF of S-N source location $p(y_s | \mathbf{D}, I)$ [lower left panel], and the marginal posterior PDF of the emission rate $p(Q | \mathbf{D}, I)$ [lower right panel] obtained from Model 3 for source reconstruction using 4 concentration detectors. All PDFs have been normalized by their maximum value p_0. The true source location is indicated using the red dot in the upper left panel.*

Table 9: *The posterior mean, maximum a posteriori (MAP) estimate, posterior standard deviation, and lower and upper bounds of the 97.5% HPD interval of the parameters x_s (m), y_s (m), and Q (g s^{-1}) obtained for Model 3 using 4 concentration detectors for source inversion. The information gain D_{KL} (measured in natural units or nits) obtained from the concentration data is summarized in the last row of the table.*

Parameter	Mean	MAP	Standard Deviation	97.5% HPD	Actual
x_s (m)	1081.5	1105.8	52.3	$(926.3, 1171.1)$	1001.7
y_s (m)	1387.5	2172.9	690.8	$(82.5, 2211.9)$	2095.7
Q (g s^{-1})	4.94	1.20	3.21	$(0.85, 10.0)$	2.00
$D_{\mathrm{KL}} = 5.6$ nits					

lower and upper 97.5% HPD intervals are tabulated in Tables 7, 8 and 9, respectively.

Note that the marginal posterior PDFs for the source parameters for Model 1 using only 4 concentration data (exhibited in Figure 14) has a much more jagged appearance than those for Model 1 shown in Figures 6 and 10 (which correspond to using 18 and 9 concentration data, respectively, for the source reconstruction). In the two latter cases, the posterior PDFs are smooth and unimodal. Even with only 4 concentration data used for the source reconstruction, the marginal posterior PDFs for the source location (x_s, y_s) in Figure 14 do peak at roughly the true location of the source. From Table 7, the source location x_s and y_s are estimated to be $\widehat{x}_s = 1058.4 \pm 59.4$ m and $\widehat{y}_s = 1988.4 \pm 395.0$ m, respectively, at one standard deviation, the true location being $(x_s, y_s) = (1001.7, 2095.7)$ m. The marginal PDFs of the source location have well-defined peaks which give MAP estimates for the x_s and y_s source locations as 1089.5 m and 2102.5 m, respectively. In this case, the MAP estimate for x_s is comparable to the posterior mean estimate for x_s, but the MAP estimate for y_s is better than that provided by the posterior mean.

Interestingly, the marginal posterior PDF for the emission rate in this case (see Figure 14) is bimodal: there is a primary peak (mode) at about 1.3 g s^{-1} and a secondary peak (mode) at about 2.0 g s^{-1} (corresponding to the true value of the emission rate). The emission rate is estimated to be 1.54 ± 0.42 g s^{-1} at one standard deviation, which is correctly estimates the actual emission rate to within two standard deviations. The MAP estimate for the emission rate in this example is 1.3 g s^{-1} and is perhaps less useful than the posterior mean, as the MAP estimate here is somewhat unrepresentative of an important portion of the posterior probability (viz., of the secondary mode that lies near 2.0 g s^{-1}).

A perusal of Figure 11 shows that the marginal posterior PDFs for the source parameters for Model 2 using 4 concentration data for the source reconstruction are smooth with well-defined peaks. In this example, the source parameters x_s, y_s and Q are estimated to be 1100.0 ± 28.7 m, 1858.8 ± 557.0 m and 2.54 ± 2.6 g s^{-1}, respectively (see Table 8). These estimates for the source parameters are quite reasonable and correspond to absolute percentage deviations from the actual values of about 9.8%, 11.3% and 30%, respectively. However, note that the uncertainties in the determination of y_s and Q are large. Although the estimate for the x_s location of the source differed from the actual x_s location by less than 10%, the estimation of the posterior standard deviation appears to be slightly too small with the result that a three standard deviation interval about the posterior mean for x_s does not contain the true value for x_s. It appears that for this example, the crude estimates s_J we used for the uncertainties σ_J are too small generally, and the distribution (inverse gamma) for the true (but unknown) uncertainties did not fully compensate for this under-estimation. This led to a small under-estimation of the posterior uncertainty for x_s, although the posterior uncertainties for y_s and Q appear to be adequate.

For Model 3 applied to the case of 4 concentration data, the marginal posterior PDFs for the source parameters are quite similar to those obtained for Model 2 (cf. Figures 15 and 16). However, we note that the marginal posterior PDFs for Model 3 are "wider" than those for Model 2, implying a greater uncertainty in the determination of the source parameters. Indeed, the source location (x_s, y_s) and emission rate Q were estimated from Model 3 to be

$\widehat{x}_s = 1081.5 \pm 52.3$ m, $\widehat{y}_s = 1387.5 \pm 690.8$ m and $\widehat{Q} = 4.94 \pm 3.21$ g s^{-1}, respectively (cf. Table 9). The precision estimates are at one standard deviation. Note that the uncertainty in the determination of y_s and Q are very large in this example, implying the information in the problem only serves to weakly constrain these two parameters. Furthermore, the large uncertainty in the determination of the location of the source in the alongwind direction (which is aligned approximately with the y_s coordinate axis since the prevailing winds were southerly) impacts the determination of the emission rate. In consequence, the emission rate is poorly determined in this example. With very few concentration data available for the source reconstruction, information about the expected uncertainty in this data is important and can result in significantly improved estimates for the source parameters (e.g., cf. Figures 14 and 15 with Figure 16). In the comparison of the three models for the posterior distribution, the parameter estimates obtained from Model 1 (which assumes that accurate or very good estimates for the uncertainty σ_J are available) should be considered as a lower bound on the estimated uncertainties. The actual parameter estimates obtained from any given concentration data set will essentially never be better than these estimates, and will almost certainly be worse when Models 2 and 3 are used (in the case where there are only crude or no estimates available for σ_J).

The information gain provided by the concentration data $\mathbf{D}$ was found to be $D_{\mathrm{KL}} = 10.1$ nits for Model 1 (cf. Table 7), but only 5.6 nits for Model 3 (cf. Table 9). Viewed in another way, the information available for the source reconstruction for Model 1 (embodied in the good estimates for the uncertainty σ_J) provided 4.5 nits of additional information over that available in Model 3, and this information allowed the "posterior volume" in the hypothesis space for Model 1 to be reduced by a factor of $\exp(4.5) \approx 90$ relative to the "posterior volume" in the hypothesis space for Model 3. Note that even a crude estimate for σ_J used in Model 2 leads to an information gain of almost 4 nits relative to that in Model 3.

7.2   European Tracer Experiment

The European Tracer Experiment (ETEX) was a major field study designed to test the predictive accuracy of models for simulating the long-range transport and dispersion of a pollutant for real-time application [35]. The dispersion experiment was conducted on 23 October 1994 and involved the release of per-fluoromethyl-cyclohexane (PMCH) as the tracer. The release site for the experiment was Monterfil in Brittany, France which is located at the following geodetic coordinates: 48.058° N and −2.0083° E. The tracer was released at a constant rate of $Q = 28.73$ kg h^{-1} over a duration of 11.83 h with the start of the release occurring at 16:00 UTC 23 October 1994 (T_b) and the end of the release occurring at 03:50 UTC 24 October 1994 (T_e). In the release, a strong west to southwesterly flow (driven on a synoptic scale by a cold front over central Europe) was advecting the tracer from the release site towards the network of samplers. PMCH concentrations were measured at 168 ground-based sampling sites located in 17 European countries. Air samplers at each of these sites monitored the concentration for 90 h after the start of the release with each measured concentration sample corresponding to a 3-h averaging time. Typically, the sampling was initiated just before the expected arrival time of the tracer cloud at each station. A total of 5,040 concentration samples were obtained.
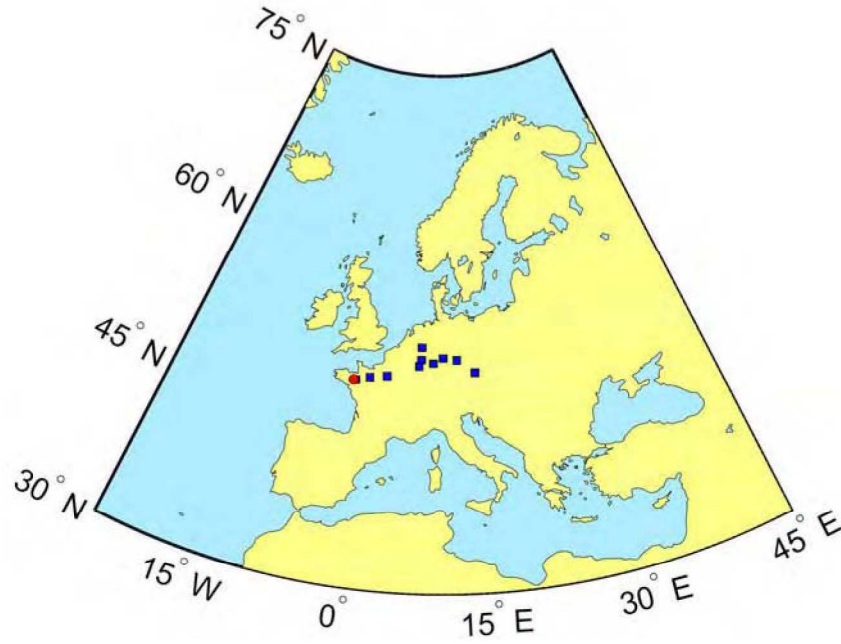
Figure 17: *Locations of the 10 sampling stations (shown by the filled blue squares) from ETEX used for the source reconstruction. The release location of the PMCH tracer source was at geodetic coordinates of 48.058° N and −2.0083° E, which was approximately 35 km west of Rennes, at Monterfil, in Brittany, France (demarcated by the filled red circle).*

For the purposes of source reconstruction, we used 35 of these concentration samples extracted from 10 sampling sites located in France, Germany and the Czech Republic with the following station codes: F02, F19, F21, D10, D13, D19, D34, D44, D45, and CR04. The locations of these 10 sampling sites relative to the location of the source are shown in Figure 17. The meteorological fields required to determine C^* were obtained from the Global Environmental Multiscale (GEM) model [36] executed in a regional configuration with a core resolution of 0.14° over Europe. The GEM model produced a series of 3 h and 6 h forecasts over the period of time corresponding to the ETEX release, with the initialization of the model coming from the Canadian Meteorological Centre (CMC) global data assimilation system. The C^* fields used in the Bayesian inference methodology for source reconstruction were computed on a polar stereographic 229×229 grid over Europe (including the United Kingdom) with a 15-km mesh length. These C^* fields were obtained using the backward Lagrangian stochastic model given by Eq. (7), which was driven with the input meteorology provided by the GEM model.

For this example, the unknown source is treated as a transient source that is described by 5 parameters; namely, $\Theta = (x_s, y_s, T_b, T_e, Q)$ where x_s and y_s are the geodetic coordinates (latitude and longitude, respectively) of the source location, T_b and T_e are the activation (source-on) and deactivation (source-off) times of the source, and Q is the source emission rate. The marginal posterior PDFs of the source parameters are summarized in Figures 18, 19 and 20 for Models 1, 2 and 3, respectively. The summary statistics for x_s, y_s, T_b, T_e, and Q obtained from the marginal posterior PDFs which include the posterior mean, MAP estimate (or posterior mode), and posterior standard deviation are tabulated in Tables 10,
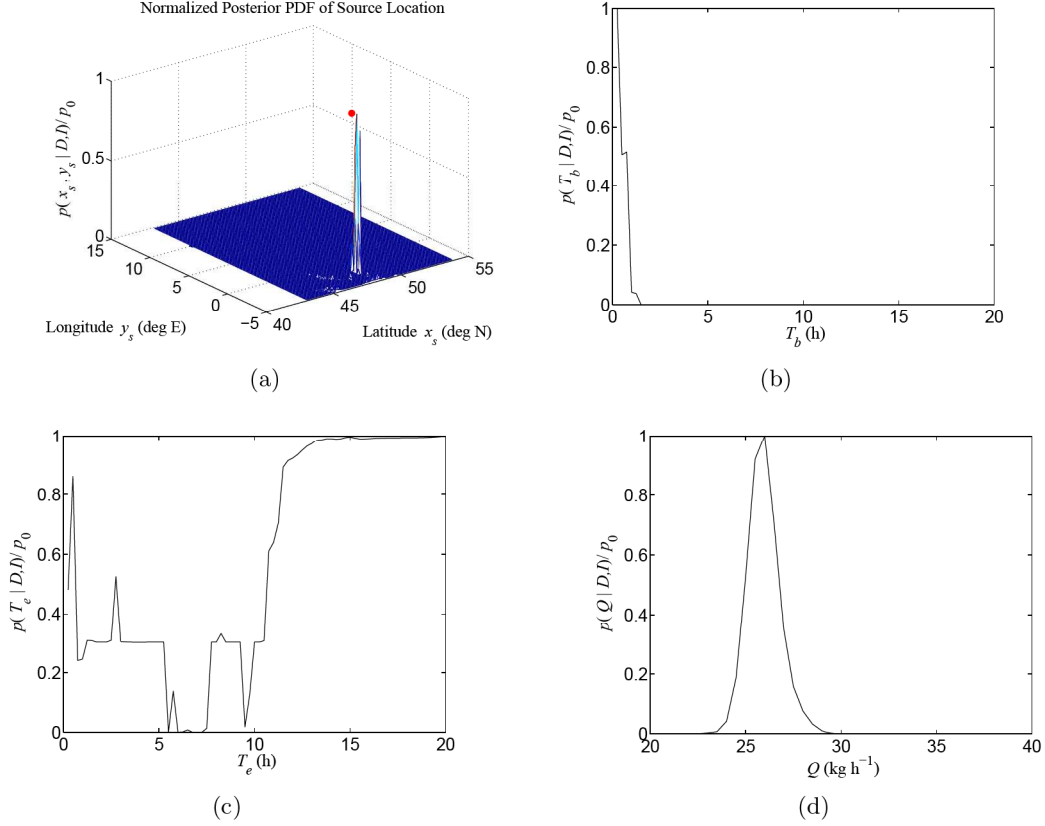
Figure 18: *The marginal joint posterior PDF of source location $p(x_s, y_s | \mathbf{D}, I)$ [upper left panel], the marginal posterior PDF of the source-on (activation) time $p(T_b | \mathbf{D}, I)$ [upper right panel], the marginal posterior PDF of the source-off (deactivation) time $p(T_e | \mathbf{D}, I)$ [lower left panel], and the marginal posterior PDF of the emission rate $p(Q | \mathbf{D}, I)$ [lower right panel] obtained from Model 1 for source reconstruction using 35 concentration data from ETEX. All PDFs have been normalized by their maximum value p_0. The true source location is indicated using the red dot in the upper left panel.*

Table 10: *The posterior mean, maximum a posteriori (MAP) estimate, and posterior standard deviation of the parameters x_s ($^\circ$ N), y_s ($^\circ$ E), T_b (h), T_e (h), and Q (kg h^{-1}) obtained for Model 1 using 35 concentration data from ETEX. The activation (T_b) and deactivation (T_e) times are referenced relative to an arbitrary time origin.*

Parameter	Mean	MAP	Standard Deviation	Actual
x_s ($^\circ$ N)	47.95	48.29	0.91	48.058
y_s ($^\circ$ E)	-2.83	-2.63	0.94	-2.0083
T_b (h)	0.47	0.25	0.26	1.0
T_e (h)	13.06	20.0	5.39	12.83
Q (kg h^{-1})	25.95	26.0	0.855	28.73
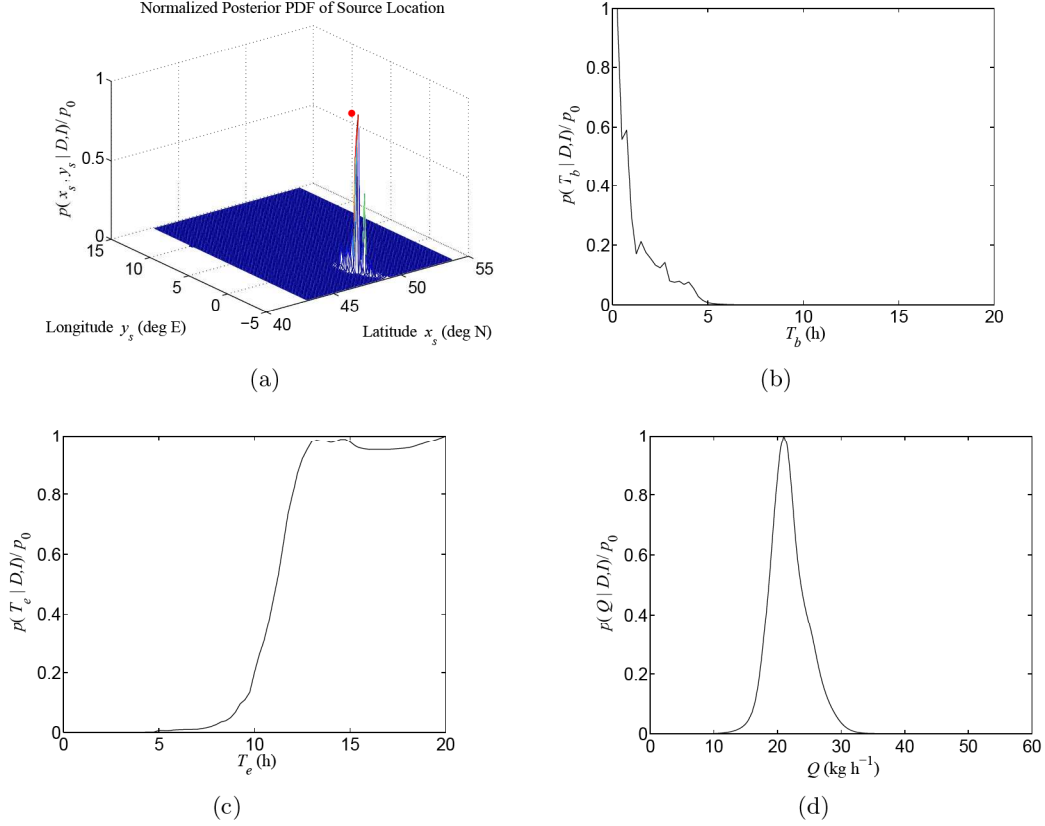
Figure 19: The marginal joint posterior PDF of source location $p(x_s, y_s | \mathbf{D}, I)$ [upper left panel], the marginal posterior PDF of the source-on (activation) time $p(T_b | \mathbf{D}, I)$ [upper right panel], the marginal posterior PDF of the source-off (deactivation) time $p(T_e | \mathbf{D}, I)$ [lower left panel], and the marginal posterior PDF of the emission rate $p(Q | \mathbf{D}, I)$ [lower right panel] obtained from Model 2 for source reconstruction using 35 concentration data from ETEX. All PDFs have been normalized by their maximum value p_0. The true source location is indicated using the red dot in the upper left panel.

Table 11: The posterior mean, maximum a posteriori (MAP) estimate, and posterior standard deviation of the parameters x_s (° N), y_s (° E), T_b (h), T_e (h), and Q (kg h^{-1}) obtained for Model 2 using 35 concentration data from ETEX. The activation (T_b) and deactivation (T_e) times are referenced relative to an arbitrary time origin.

| Parameter	Mean	MAP	Standard Deviation	Actual
x_s (° N)	48.20	48.17	0.21	48.058
y_s (° E)	−2.48	−2.59	0.72	−2.0083
T_b (h)	1.30	0.25	1.17	1.0
T_e (h)	15.42	20.0	2.83	12.83
Q (kg h^{-1})	21.81	21.0	3.05	28.73
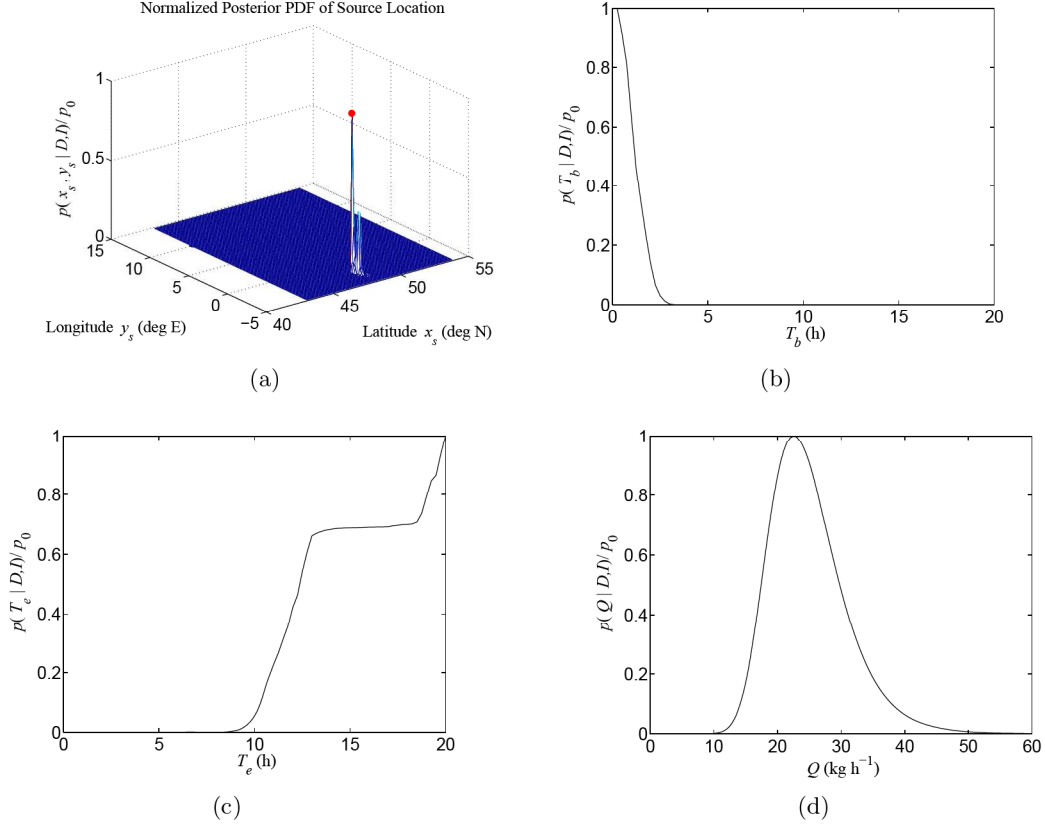
Figure 20: *The marginal joint posterior PDF of source location $p(x_s, y_s|\mathbf{D}, I)$ [upper left panel], the marginal posterior PDF of the source-on (activation) time $p(T_b|\mathbf{D}, I)$ [upper right panel], the marginal posterior PDF of the source-off (deactivation) time $p(T_e|\mathbf{D}, I)$ [lower left panel], and the marginal posterior PDF of the emission rate $p(Q|\mathbf{D}, I)$ [lower right panel] obtained from Model 3 for source reconstruction using 35 concentration data from ETEX. All PDFs have been normalized by their maximum value p_0. The true source location is indicated using the red dot in the upper left panel.*

Table 12: *The posterior mean, maximum a posteriori (MAP) estimate, and posterior standard deviation of the parameters x_s ($^\circ$ N), y_s ($^\circ$ E), T_b (h), T_e (h), and Q (kg h^{-1}) obtained for Model 3 using 35 concentration data from ETEX. The activation (T_b) and deactivation (T_e) times are referenced relative to an arbitrary time origin.*

Parameter	Mean	MAP	Standard Deviation	Actual
x_s ($^\circ$ N)	48.15	48.13	0.058	48.058
y_s ($^\circ$ E)	-2.36	-1.99	0.38	-2.0083
T_b (h)	0.86	0.25	0.54	1.0
T_e (h)	15.94	20.0	2.65	12.83
Q (kg h^{-1})	24.96	22.5	6.08	28.73

11 and 12, respectively.

For the case of Model 1 (see Figure 18 and Table 10), the location of the source was estimated to be at latitudinal and longitudinal coordinates of $47.95° \pm 0.91°$ N and $-2.83° \pm 0.94°$ E, respectively; and, the emission rate was estimated to be 25.95 ± 0.855 kg h^{-1} (with precision estimates at one standard deviation). The actual location of the source was correctly estimated to within a one standard deviation interval; and, the emission rate was correctly determined to within a three standard deviation interval. The (marginal) posterior distributions of the source activation T_b and deactivation T_e times yielded the following estimates: 0.468 ± 0.257 h and 13.06 ± 5.39 h, respectively. These times are referenced with respect to an arbitrary time origin, which in this case was chosen to be one hour before the start of the release (the latter of which occurred at 16:00 UTC 23 October 1994). Note that the concentration data used for the source determination here only weakly constrains the source deactivation time which was estimated with a large uncertainty. In particular, a perusal of the posterior PDF of T_e [cf. Figure 18(c)] shows that it has a multimodal form — this feature seems to suggest that the model and measurement uncertainty (encoded in σ_J for Model 1) has been under-estimated in this case, causing the reconstruction to interpret "noise" as signal and resulting in the possible multimodal form for $p(T_e|\mathbf{D}, I)$.

In view of the fact that the actual uncertainties appear to have been under-estimated, the source reconstruction using Model 2 was undertaken with the scale parameter α set to $\alpha = 2$. Recall that specification $\alpha > 1$ codes for an expected negative bias in the user's estimates for the actual uncertainties (viz., the user has under-estimated the magnitude of the true model and measurement uncertainties on average). An examination of the posterior distributions of the source parameters obtained using Model 2 (cf. Figure 19) shows that these distributions tend to be broader than those obtained using Model 1 (cf. Figure 18). For Model 2, the best estimates (and uncertainties at one standard deviation) of the source parameters based on the posterior mean are as follows: $\widehat{x}_s = 48.20° \pm 0.21°$ N, $\widehat{y}_s = -2.48° \pm 0.72°$ E, $\widehat{T}_b = 1.30 \pm 1.17$ h, $\widehat{T}_e = 15.42 \pm 2.83$ h, and $\widehat{Q} = 21.81 \pm 3.05$ kg h^{-1}. Note that the marginal distribution for T_e is much smoother for Model 2 — again, it is seen that the information contained in the concentration data do not allow the deactivation time of the source to be determined with any degree of precision. Nevertheless, note that $p(T_e|\mathbf{D}, I)$ increases sharply at about $T_e = 12.5$ h suggesting that the probability that the source was turned off before about 12.5 h (relative to the arbitrary time origin) is small. However, there is a very large probability that the source was turned off after about 12.5 h, although the concentration data do not allow a good estimate to be obtained for T_e.

Finally, the source reconstruction using Model 3 provides the most conservative recovery of the source parameters (cf. Figure 20 and Table 12). The analysis based on Model 3 yielded the following source parameter estimates expressed as posterior mean values and their uncertainties at one standard deviation: $\widehat{x}_s = 48.15° \pm 0.06°$ N, $\widehat{y}_s = -2.36° \pm 0.38°$ E, $\widehat{T}_b = 0.86 \pm 0.54$ h, $\widehat{T}_e = 15.94 \pm 2.65$ h, and $\widehat{Q} = 24.96 \pm 6.08$ kg h^{-1} As can be seen, the recovery of the parameters for the source is very good. All the source parameters have been correctly determined to within a one standard deviation interval (approximately or better).

8 Conclusions

In this report, a technical description of urbanSOURCE is provided. The module urbanSOURCE is an operational implementation of a sensor-driven modeling paradigm for source reconstruction: namely, the determination of the parameters that define an unknown source, given a finite number of noisy concentration measurements obtained from a network of CBRN sensors. To this purpose, we have formulated a Bayesian inferential scheme for the joint determination of the parameters of an unknown source. In particular, three different model equations have been formulated for the likelihood function leading to three different models for the posterior PDF of the source parameters. These models reflect three different states of knowledge regarding the uncertainties in the concentration (measured and predicted) used for the source reconstruction (which must necessarily include the contributions to the expected discrepancy between the measured concentration d_J and predicted concentration $\overline{C}_J$ arising from the effects of the input, model, stochastic and measurement errors).

We have illustrated the application of urbanSOURCE to source reconstruction in a complex environment; namely, for contaminant transport and dispersion in an urban environment (JU2003) and in a complex terrain environment (ETEX) involving highly disturbed mean wind and turbulence fields exhibiting a significant degree of spatial inhomogeneity and/or temporal non-stationarity (viz., unsteadiness in the mean wind and turbulence). To this purpose, the posterior PDFs of the source parameters for a localized (idealized here as a point) source were inferred using actual measured concentration data from the JU2003 field experiment in Oklahoma City and from the European Tracer Experiment over continental Europe.

The methodology implemented in urbanSOURCE has been successfully applied to estimate the parameters for a continuous source and their associated uncertainties for three different cases corresponding to a particular field trial in JU2003. These three cases involved the use of 18, 9 and 4 concentration data in the source reconstruction. In addition, the methodology was applied to reconstruction of the parameters of a transient source using 35 concentration data obtained from 10 sampling sites in ETEX. For each of these examples/cases, the reconstruction was undertaken using the three different models for the posterior distribution of the source parameters. It is shown that the parameters (e.g., location, emission rate, source-on and source-off times) that characterize the source (either continuous or transient) have been recovered correctly. Finally, the methodology provides a rigorous determination of the uncertainty (e.g., standard deviation, credible intervals) in the inference of the source parameters (allowing both the accuracy and precision in the source parameter estimation to be assessed).

The next step is to integrate urbanSOURCE as an operational capability for source reconstruction into the integrative multiscale urban modeling system implemented in the computational infrastructure at a government operations facility (Environmental Emergency Response Section at Canadian Meteorological Centre). This will involve interfacing urbanSOURCE with the modules urbanAEU and urbanBLS that are used to compute the

adjunct concentration field C^*, required for the rapid calculation of the likelihood function in urbanSOURCE. Finally, to complete the sensor-driven modeling paradigm, the capability needs to be interfaced with information (warning and reporting) systems for automated data acquisition from CBRN sensors in the field of operations. This linkage will allow the mutual optimization of CBRN sensor data and models, enabling the rapid estimation of unknown source terms from sensor data followed by an accurate prediction of the transport, dispersion and fate of the toxic agent.

References

[1] Stern, A. C., Boubel, R. W., Turner, D. B., and Fox, D. L. (1983), *Fundamentals of Air Pollution*, 2nd Ed., Academic Press Inc., New York.

[2] United States Government Accountability Office (2008), First Responders' Ability to Detect and Model Hazardous Releases in Urban Areas Is Significantly Limited (online), http://www.gao.gov/new.items/d08180.pdf (Access Date: August 2008). Report to Congressional Requesters, GAO–08–180.

[3] Yee, E., Lien, F.-S., and Ji, H. (2007), Technical description of urban microscale modeling system: Component 1 of CRTI Project 02-0093RD, (DRDC Suffield TR 2007-067) Defence R&D Canada – Suffield.

[4] Lien, F.-S., Ji, H., and Yee, E. (2007), Parallelization of an urban microscale flow model (urbanSTREAM): Component 1 of CRTI Project 02-0093RD, (WATCFD Report WD001/07) Waterloo Engineering CFD Consulting, Inc., Waterloo, Ontario.

[5] Wilson, J.D. (2007), Technical description of urban microscale modeling system: Component 4 of CRTI Project 02-0093RD, (Report JDW01/07) J.D. Wilson and Associates, Edmonton, Alberta.

[6] Yee, E. and Hogue, R. (2008), Advanced integrative multiscale modeling system for countering the threat of CBRN terrorism, In Ruiz, Juan J., (Ed.), *Proceedings of NATO Modeling and Simulation Group Conference (MSG-060)*.

[7] Shea, D. A. and Lister, S. A. (2003), The BioWatch Program: Detection of Bioterrorism (online), http://www.fas.org/sgp/crs/terror/RL32152.html (Access Date: September 2008). Congressional Research Service Report No. RL 35152.

[8] Van der Vink, G. E. and Park, J. (1994), Nuclear test ban monitoring: new requirements, new resources, *Science*, 263, 634–635.

[9] Hourdin, F. and Issartel, J.-P. (2000), Sub-surface nuclear tests monitoring through the CTBT xenon network, *Geophysical Research Letters*, 27, 2245–2248.

[10] Keats, A. (2009), *Bayesian inference for source determination in the atmospheric environment*, Ph.D Thesis, University of Waterloo, Waterloo, Ontario.

[11] Yee, E. (2005), Probabilistic inference: an application to the inverse problem of source function estimation, In *The Technical Cooperation Program (TTCP) Chemical and Biological Defence (CBD) Group Technical Panel 9 (TP-9) Annual Meeting*. Defence Science and Technology Organization, Melbourne, Australia.

[12] Yee, E. (2006), A Bayesian approach for reconstruction of the characteristics of a localized pollutant source from a small number of concentration measurements obtained by spatially distributed "electronic noses", In *Russian-Canadian Workshop on Modeling of Atmospheric Dispersion of Weapon Agents*. Karpov Institute of Physical Chemistry, Moscow, Russia.

[13] Yee, E. (2007), Bayesian probabilistic approach for inverse source determination from limited and noisy chemical or biological sensor concentration measurements, In Fountain III, Augustus W., (Ed.), *Proceedings of SPIE, Chemical and Biological Sensing VIII*, Vol. 6554, 65540W, doi:10.1117/12.721630.

[14] Yee, E. (2008), Inverse dispersion of an unknown number of contaminant sources, In *Proceedings of 15th Joint Conference on the Applications of Air Pollution Meteorology with the A&WMA*. New Orleans, LA, Paper 7.1.

[15] Yee, E. (2008), Theory for reconstruction of an unknown number of contaminant sources using probabilistic inference, *Boundary-Layer Meteorology*, 127, 359–394.

[16] Yee, E. (2009), Validation of a sensor-driven modeling paradigm for multiple source reconstruction with FFT-07 data, (DRDC Suffield TR 2009-040) Defence R&D Canada – Suffield.

[17] Yee, E., Lien, F.-S., Keats, A., Hseih, K.J., and D'Amours, R. (2006), Validation of Bayesian inference for emission source distribution using the Joint Urban 2003 and European Tracer Experiments, In *Fourth International Symposium on Computational Wind Engineering (CWE2006)*. Yokohama, Japan.

[18] Yee, E., Lien, F.-S., Keats, A., and D'Amours, R. (2008), Bayesian inversion of concentration data: Source reconstruction in the adjoint representation of atmospheric diffusion, *Journal of Wind Engineering and Industrial Aerodynamics*, 96, 1805–1816.

[19] Keats, A., Yee, E., and Lien, F.-S. (2007), Bayesian inference for source determination with applications to a complex urban environment, *Atmospheric Environment*, 41, 465–479.

[20] Keats, A., Yee, E., and Lien, F.-S. (2007), Efficiently characterizing the origin and decay rate of a nonconservative scalar using probability theory, *Ecological Modeling*, 205, 437–452.

[21] Thomson, D. J. (1987), Criteria for the selection of stochastic models of particle trajectories in turbulent flows, *Journal of Fluid Mechanics*, 180, 529–556.

[22] Flesch, T., Wilson, J. D., and Yee, E. (1995), Backward-time Lagrangian stochastic dispersion models and their application to estimate gaseous emissions, *Journal of Applied Meteorology*, 34, 1320–1332.

[23] Rao, K. S. (2005), Uncertainty analysis in atmospheric dispersion modeling, *Pure and Applied Geophysics*, 162, 1893–1917.

[24] Yee, E., Chan, R., Kosteniuk, P. R., Chandler, G. M., Biltoft, C. A., and Bowers, J. F. (1994), Experimental measurements of concentration fluctuations and scales in a dispersing plume in the atmospheric surface layer obtained using a very fast-response concentration detector, *Journal of Applied Meteorology*, 33, 996–1016.

[25] Yee, E., Chan, R., Kosteniuk, P. R., Chandler, G. M., Biltoft, C. A., and Bowers, J. F. (1995), The vertical structure of concentration fluctuation statistics in plumes dispersing in the atmospheric surface layer, *Boundary-Layer Meteorology*, 76, 41–67.

[26] Yee, E. and Biltoft, C. A. (2004), Concentration fluctuation measurements in a plume dispersing through a regular array of obstacles, *Boundary-Layer Meteorology*, 111, 363–415.

[27] Cox, R. T. (1946), Probability, frequency, and reasonable expectation, *American Journal of Physics*, 14, 1–13.

[28] Jaynes, E. T. (2003), *Probability Theory: The Logic of Science*, Cambridge University Press, Cambridge, UK.

[29] Senocak, I., Hengartner, N. W., Short, M. B., and Daniel, W. B. (2008), Stochastic event reconstruction of atmospheric contaminant dispersion using Bayesian inference, *Atmospheric Environment*, 42, 7718–7727.

[30] Jeffreys, H. (1961), *Theory of Probability*, Oxford University Press, London.

[31] Thomson, L. C., Hirst, B., Gibson, G., Gillespie, S., Jonathan, P., Skeldon, K. D., and Padgett, M. J. (2007), An improved algorithm for locating a gas source using inverse methods, *Atmospheric Environment*, 41, 1128–1134.

[32] Allen, C. T., Young, G., and Haupt, S. E. (2007), Improving pollutant source characterization by better estimating wind direction with a genetic algorithm, *Atmospheric Environment*, 41, 2283–2289.

[33] Cover, T. M. and Thomas, J. A. (1991), *Elements of Information Theory*, John Wiley & Sons, Inc., New York.

[34] Allwine, K. J., Leach, M. J., Stockham, L. W., Shinn, J. S., Hosker, R. P., Bowers, J. F., and Pace, J. C. (2004), Overview of Joint Urban 2003–an atmospheric dispersion study in Oklahoma City, In *Symposium on Planning, Nowcasting, and Forecasting in the Urban Zone*. American Meteorological Society, Seattle, Washington.

[35] Nodop, K., Connolly, R., and Girardi, F. (1998), The field campaigns of the European Tracer Experiment (ETEX): overview and results, *Atmospheric Environment*, 32, 4095–4108.

[36] D'Amours, R. (1998), Modeling the ETEX plume dispersion with the Canadian emergency response model, *Atmospheric Environment*, 32, 4335–4341.

DOCUMENT CONTROL DATA

(Security classification of title, body of abstract and indexing annotation must be entered when document is classified)

1. ORIGINATOR (The name and address of the organization preparing the document. Organizations for whom the document was prepared, e.g. Centre sponsoring a contractor's report, or tasking agency, are entered in section 8.) **Defence R&D Canada – Suffield** **Box 4000, Station Main, Medicine Hat, Alberta,** **Canada T1A 8K6**	2. SECURITY CLASSIFICATION (Overall security classification of the document including special warning terms if applicable.) **UNCLASSIFIED**

3. TITLE (The complete document title as indicated on the title page. Its classification should be indicated by the appropriate abbreviation (S, C or U) in parentheses after the title.) **An Operational Implementation of a CBRN Sensor-Driven Modeling Paradigm for Stochastic Event Reconstruction**

4. AUTHORS (Last name, followed by initials – ranks, titles, etc. not to be used.) **Yee, E.**

5. DATE OF PUBLICATION (Month and year of publication of document.) **May 2010**	6a. NO. OF PAGES (Total containing information. Include Annexes, Appendices, etc.) **68**	6b. NO. OF REFS (Total cited in document.) **36**

7. DESCRIPTIVE NOTES (The category of the document, e.g. technical report, technical note or memorandum. If appropriate, enter the type of report, e.g. interim, progress, summary, annual or final. Give the inclusive dates when a specific reporting period is covered.) **Technical Report**

8. SPONSORING ACTIVITY (The name of the department project office or laboratory sponsoring the research and development – include address.) **Defence R&D Canada – Suffield** **Box 4000, Station Main, Medicine Hat, Alberta, Canada T1A 8K6**

9a. PROJECT NO. (The applicable research and development project number under which the document was written. Please specify whether project or grant.) **CRTI Project 07-0196TD**	9b. GRANT OR CONTRACT NO. (If appropriate, the applicable number under which the document was written.)

10a. ORIGINATOR'S DOCUMENT NUMBER (The official document number by which the document is identified by the originating activity. This number must be unique to this document.) **DRDC Suffield TR 2010-070**	10b. OTHER DOCUMENT NO(s). (Any other numbers which may be assigned this document either by the originator or by the sponsor.)

11. DOCUMENT AVAILABILITY (Any limitations on further dissemination of the document, other than those imposed by security classification.) (X) Unlimited distribution (   ) Defence departments and defence contractors; further distribution only as approved (   ) Defence departments and Canadian defence contractors; further distribution only as approved (   ) Government departments and agencies; further distribution only as approved (   ) Defence departments; further distribution only as approved (   ) Other (please specify):

12. DOCUMENT ANNOUNCEMENT (Any limitation to the bibliographic announcement of this document. This will normally correspond to the Document Availability (11). However, where further distribution (beyond the audience specified in (11)) is possible, a wider announcement audience may be selected.) **Unlimited**

13. ABSTRACT (A brief and factual summary of the document. It may also appear elsewhere in the body of the document itself. It is highly desirable that the abstract of classified documents be unclassified. Each paragraph of the abstract shall begin with an indication of the security classification of the information in the paragraph (unless the document itself is unclassified) represented as (S), (C), (R), or (U). It is not necessary to include here abstracts in both official languages unless the text is bilingual.)

This report provides a technical description of the module urbanSOURCE, which is an operational implementation of an innovative sensor-driven modeling paradigm for source reconstruction. This module permits the rapid and robust estimation of the parameters of an unknown source, using a finite number of noisy concentration measurements obtained from a sensor array. The problem is solved using a Bayesian probabilistic inferential framework in which Bayesian probability theory is used to formulate the posterior distribution for the source parameters. Three different model equations have been formulated for the likelihood function, leading to three different models for the posterior distribution of the source parameters. The application of the methodology implemented in urbanSOURCE is illustrated using real dispersion data obtained from two examples (Joint Urban 2003 field experiment in Oklahoma City and European Tracer Experiment) involving contaminant dispersion in highly disturbed flows over urban and complex environments, where the idealizations of horizontal homogeneity and/or temporal stationarity in the flow cannot be applied to simplify the problem.

14. KEYWORDS, DESCRIPTORS or IDENTIFIERS (Technically meaningful terms or short phrases that characterize a document and could be helpful in cataloguing the document. They should be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location may also be included. If possible keywords should be selected from a published thesaurus. e.g. Thesaurus of Engineering and Scientific Terms (TEST) and that thesaurus identified. If it is not possible to select indexing terms which are Unclassified, the classification of each should be indicated as with the title.)

atmospheric dispersion
Bayesian inference
sensor arrays
sensor/model data fusion
source reconstruction